Process for preparing ketocarotenoids in genetically modified organisms

The present invention relates to a process for preparing ketocarotenoids by cultivation of genetically modified organisms which, compared with the wild type, have a modified ketolase activity, to the genetically modified organisms, and to the use thereof as human and animal foods and for producing ketocarotenoid extracts.

Carotenoids are synthesized de novo in bacteria, algae, fungi and plants. Ketocarotenoids, i.e. carotenoids containing at least one keto group, such as, for example, astaxanthin, canthaxanthin, echinenone, 3-hydroxyechinenone, 3'-hydroxyechinenone, adonirubin and adonixanthin, are natural antioxidants and pigments produced as secondary metabolites by some algae and microorganisms.

Because of their coloring properties, the ketocarotenoids and especially astaxanthin are used as pigmenting aids in livestock nutrition, especially in trout, salmon and shrimp rearing.

Astaxanthin is currently prepared for the most part by chemical synthesis processes. Natural ketocarotenoids such as, for example, natural astaxanthin are currently obtained in small quantities in biotechnological processes by cultivation of algae, for example *Haematococcus pluvialis* or by fermentation of genetically optimized microorganisms and subsequent isolation.

An economic biotechnological process for preparing natural ketocarotenoids is therefore of great importance.

Nucleic acids encoding a ketolase and the corresponding protein sequences have been isolated from various organisms and annotated, such as, for example, nucleic acids encoding a ketolase from *Agrobacterium aurantiacum* (EP 735 137, Accession No. D58420), from *Alcaligenes sp.* PC-1 (EP 735137, Accession No. D58422), *Haematococcus pluvialis Flotow em. Wille* and *Haematoccus pluvialis, NIES-144* (EP 725137, WO 98/18910 and Lotan et al, FEBS Letters 1995, 364, 125-128, Accession No. X86782 and D45881, *Paracoccus marcusii* (Accession No. Y15112), *Synechocystis sp. Strain PC6803* (Accession No. NP_442491), *Bradyrhizobium sp.* (Accession No. AF218415) and *Nostoc sp. PCC 7120* (Kaneko et al., DNA Res. 2001, 8(5), 205-213; Accession No. AP003592, BAB74888).

EP 735 137 describes the preparation of xanthophylls in microorganisms such as, for example, *E. coli* by introducing ketolase genes (crtW) from *Agrobacterium aurantiacum* or *Alcaligenes sp. PC-1* into microorganisms.

EP 725 137, WO 98/18910, Kajiwara et al. (Plant Mol. Biol. 1995, 29, 343-352) and Hirschberg et al. (FEBS Letters 1995, 364, 125-128) disclose the preparation of astaxanthin by introducing

ketolase genes from *Haematococcus pluvialis* (crtW, crtO or bkt) into *E. coli*.

Hirschberg et al. (FEBS Letters 1997, 404, 129-134) describe the preparation of astaxanthin in *Synechococcus* by introducing ketolase genes (crtO) from *Haematococcus pluvialis*. Sandmann et al. (Photochemistry and Photobiology 2001, 73(5), 551-55) describe an analogous process which, however, leads to the preparation of canthaxanthin and provides only traces of astaxanthin.

WO 98/18910 and Hirschberg et al. (Nature Biotechnology 2000, 18(8), 888-892) describe the synthesis of ketocarotenoids in nectaries of tobacco flowers by introducing the ketolase gene from *Haematococcus pluvialis* (crtO) into tobacco.

WO 01/20011 describes a DNA construct for producing ketocarotenoids, especially astaxanthin, in seeds of oilseed crops such as rape, sunflower, soybean and mustard, using a seed-specific promoter and a ketolase from *Haematococcus pluvialis*.

All the processes described in the prior art for preparing ketocarotenoids and, in particular, the processes described for preparing astaxanthin have the disadvantage that the transgenic organisms provide a large quantity of hydroxylated byproducts, such as zeaxanthin and adonixanthin, for example.

It is an object of the present invention to provide a process for preparing ketocarotenoids by cultivation of genetically modified organisms, and to provide further genetically modified organisms which produce ketocarotenoids, which have the prior art disadvantages described above to a smaller extent or not at all.

We have found that this object is achieved by a process for preparing ketocarotenoids by cultivating genetically modified organisms which, compared with the wild type, have a modified ketolase activity, and the modified ketolase activity is caused by a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2.

The organisms of the invention, such as, for example, microorganisms or plants, are preferably able as starting organisms naturally to produce carotenoids such as, for example, β-carotene or zeaxanthin, or can be made able by genetic modification such as, for example, reregulation of metabolic pathways or complementation to produce carotenoids such as, for example, β-carotene or zeaxanthin.

Some organisms are already able as starting or wild-type organisms to produce ketocarotenoids such as, for example, astaxanthin or canthaxanthin. These organisms, such as, for example, *Haematococcus pluvialis*, *Paracoccus marcusii*, *Xanthophyllomyces dendrorhous*, *Bacillus circulans*, *Chlorococcum*, *Phaffia rhodozyma*, *adonis*, *Neochloris wimmeri*, *Protosiphon botryoides*, *Scotiellopsis oocystiformis*, *Scenedesmus vacuolatus*, *Chlorela zofingiensis*, *Ankistrodesmus braunii*, *Euglena sanguinea*, *Bacillus atrophaeus*, Blakeslea already have as starting or wild-type organism a ketolase activity.

In one embodiment of the process of the invention, therefore, the starting organisms used are those already having a ketolase activity as wild type or starting organism. In this embodiment, the genetic modification brings about an increase in the ketolase activity compared with the wild type or starting organism.

Ketolase activity means the enzymic activity of a ketolase.
A ketolase means a protein which has the enzymatic activity of introducing a keto group on the, optionally substituted, β-ionone ring of carotenoids.

A ketolase means in particular a protein having the enzymatic activity of converting β-carotene into canthaxanthin.

Accordingly, ketolase activity means the amount of β-carotene converted or amount of canthaxanthin produced in a particular time by the ketolase protein.

Thus, when a ketolase activity is increased compared with the wild type, the amount of β-carotene converted or the amount of canthaxanthin produced in a particular time is increased by the ketolase protein compared with the wild type.

This increase in the ketolase activity is preferably at least 5%, more preferably at least 20%, more preferably at least 50%, more preferably at least 100%, preferably at least 300%, more preferably at least 500%, in particular at least 600%, of the ketolase activity of the wild type.

The term "wild type" means according to the invention the corresponding starting organism.

Depending on the context, the term "organism" may mean the starting organism (wild type) or a genetically modified organism of the invention, or both.

"Wild type" means, preferably and especially in cases where the organism or the wild type cannot be unambiguously assigned, in each case a reference organism for the increasing or causing of the ketolase activity, for the increasing, described hereinafter, of the hydroxylase

activity, for the increasing, described hereinafter, of the β-cyclase activity and the increasing of the content of ketocarotenoids.

This reference organism for microorganisms which already have a ketolase activity as wild type is preferably *Haematococcus pluvialis*.

This reference organism for microorganisms which have no ketolase activity as wild type is preferably Blakeslea.

This reference organism for plants which already have a ketolase activity as wild type is preferably *Adonis aestivalis, Adonis flammeus* or *Adonis annuus*, particularly preferably *Adonis aestivalis*.

This reference organism for plants which have no ketolase activity in petals as wild type is preferably *Tagetes erecta, Tagetes patula, Tagetes lucida, Tagetes pringlei, Tagetes palmeri, Tagetes minuta* or *Tagetes campanulata*, particularly preferably *Tagetes erecta*.

Determination of the ketolase activity in the genetically modified organisms of the invention and in wild-type and reference organisms preferably takes place under the following conditions:

Determination of the ketolase activity in plant or microorganism material is based on the method of Frazer et al., (J. Biol. Chem. 272(10): 6128-6135, 1997). The ketolase activity in plant or microorganism extracts is determined using the substrates β-carotene and canthaxanthin in the presence of lipid (soybean lecithin) and detergent (sodium cholate). Substrate/product ratios from ketolase assays are measured by means of HPLC.

Various ways are possible for increasing the ketolase activity, for example by switching off inhibitory regulatory mechanisms at the translation and protein level or by increasing the gene expression of a nucleic acid encoding a ketolase compared with the wild type, for example by inducing the ketolase gene by activators or by introducing nucleic acids encoding a ketolase into the organism.

Increasing the gene expression of a nucleic acid encoding a ketolase also means according to the invention in this embodiment the manipulation of the expression of the organisms own endogenous ketolases. This can be achieved for example by modifying the promoter DNA sequence for ketolase-encoding genes. Such a modification, which results in a modified or, preferably, increased expression rate of at least one endogenous ketolase gene, can also be effected by deletion or insertion of DNA sequences.

It is possible as described above to modify the expression of at least one endogenous ketolase through application of exogenous stimuli. This can be effected by particular physiological conditions, i.e. through application of foreign substances.

A further possibility for achieving an increased expression of at least one endogenous ketolase gene is for a regulator protein which does not occur in the wild-type organism or is modified to interact with the promoter of these genes.

A regulator of this type may be a chimeric protein which consists of a DNA-binding domain and of a transcription activator domain as described, for example, in WO 96/06166.

In a preferred embodiment, the ketolase activity is increased by comparison with the wild type by increasing the gene expression of a nucleic acid encoding a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2.

In a further preferred embodiment, the gene expression of a nucleic acid encoding a ketolase is increased by introducing nucleic acids which encode ketolases, where the ketolases have the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2, into the organisms.

Thus, in this embodiment, at least one further ketolase gene encoding a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2, is present in the transgenic organisms of the invention compared with the wild type.

In this embodiment, the genetically modified organism of the invention accordingly has at least one exogenous (= heterologous) nucleic acid encoding a ketolase, or has at least two endogenous nucleic acids encoding a ketolase, where the ketolases comprise the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2.

In another, preferred embodiment of the process of the invention, the organisms used as starting organisms have no ketolase activity as wild type.

In this preferred embodiment, the genetic modification causes the ketolase activity in the organisms. The genetically modified organism of the invention thus has in this preferred embodiment a ketolase activity compared with the genetically unmodified wild type, and is thus preferably capable of transgenic expression of a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2.

In this preferred embodiment, the gene expression of a nucleic acid encoding a ketolase is caused, in analogy to the increasing, described above, of the gene expression of a nucleic acid encoding a ketolase, preferably by introducing nucleic acids which encode ketolases comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2, into the starting organism.

It is possible to use for this purpose in both embodiments in principle all nucleic acids which encode a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2.

The use of the nucleic acids of the invention encoding a ketolase leads in the process of the invention surprisingly to ketocarotenoids having a smaller quantity of hydroxylated byproducts than on use of the ketolase genes used in the prior art.

All the nucleic acids mentioned in the description may be, for example, an RNA, DNA or cDNA sequence.

In the case of genomic ketolase sequences from eukaryotic sources which comprise introns, it is preferred to use nucleic acid sequences which have already been processed, such as the corresponding cDNAs, in the case where the host organism is unable or cannot be made able to express the corresponding ketolase.

Examples of nucleic acids encoding a ketolase, and the corresponding ketolases comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2, which can be used advantageously in the process of the invention are, for example, sequences from

Nostoc sp. Strain PCC7120 (Accession No. AP003592, BAB74888; nucleic acid: SEQ ID NO: 1,

protein: SEQ ID NO: 2),

Nostoc punctiforme ATTC 29133, nucleic acid: Acc. No. NZ_AABC01000195, base pairs 55,604 to 55,392 (SEQ ID NO: 3); protein: Acc. No. ZP_00111258 (SEQ ID NO: 4) (annotated as a putative protein) or

Nostoc punctiforme ATTC 29133, nucleic acid: Acc. No. NZ_AABC01000196, base pairs 140,571 to 139,810 (SEQ ID NO: 5), protein: (SEQ ID NO: 6) (not annotated)

Synechococcus sp. WH 8102, nucleic acid: Acc. No. NZ_AABD01000001, base pairs 1,354,725-1,355,528 (SEQ ID NO: 46), protein: Acc. No. ZP_00115639 (SEQ ID NO: 47) (annotated as a putative protein),

Nodularia spumigena NSOR10, (Accession No. AY210783, AAO64399; nucleic acid: SEQ ID NO: 52, protein: SEQ ID NO: 53)

or ketolase sequences derived from said sequences, such as, for example,

the ketolases of the sequence SEQ ID NO: 8 or 10 and the corresponding coding nucleic acid sequences SEQ ID NO: 7 or SEQ ID NO: 9 which arise, for example, from variation/mutation of the sequences SEQ ID NO: 4 and SEQ ID NO: 3, respectively,

the ketolases of the sequence SEQ ID NO: 12 or 14 and the corresponding coding nucleic acid sequences SEQ ID NO: 11 or SEQ ID NO: 13 which arise, for example, from variation/mutation of the sequences SEQ ID NO: 6 and SEQ ID NO: 5, respectively,

the ketolases of the sequence SEQ ID NO: 49 or 51 and the corresponding coding nucleic acid sequences SEQ ID NO: 48 or SEQ ID NO: 50 which arise, for example, from variation and mutation of the sequences SEQ ID NO: 47 and SEQ ID NO: 46, respectively,

Further natural examples of ketolases and ketolase genes which can be used in the process of the invention can easily be found for example from various organisms whose genomic sequence is known through identity comparisons of the amino acid sequences or of the corresponding back-translated nucleic acid sequences from databases with the sequence SEQ ID NO: 2 described above.

Further natural examples of ketolases and ketolase genes can additionally be easily found starting from the nucleic acid sequences above, in particular starting from the sequence SEQ ID NO: 1 from various organisms whose genomic sequence is unknown through hybridization

techniques in a manner known per se.

The hybridization can take place under moderate (low stringency) or preferably under stringent (high stringency) conditions.

Hybridization conditions of these types are described for example in Sambrook, J., Fritsch, E.F., Maniatis, T., in: Molecular Cloning (A Laboratory Manual), 2nd edition, Cold Spring Harbor Laboratory Press, 1989, pages 9.31-9.57 or in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.

For example, the conditions during the washing step can be selected from the range of conditions limited by those of low stringency (with 2X SSC at 50°C) and those of high stringency (with 0.2X SSC at 50°C, preferably at 65°C) (20X SSC: 0.3 M sodium citrate, 3 M sodium chloride, pH 7.0).

An additional possibility is to raise the temperature during the washing step from moderate conditions at room temperature, 22°C, up to stringent conditions at 65°C.

Both parameters, the salt concentration and temperature, can be varied simultaneously, and it is also possible to keep one of the two parameters constant and vary only the other one. It is also possible to employ denaturing agents such as, for example, formamide or SDS during the hybridization. Hybridization in the presence of 50% formamide is preferably carried out at 42°C.

Some examples of conditions for hybridization and washing step are given below:

(1) Hybridization conditions with for example

(i) 4X SSC at 65°C, or

(ii) 6X SSC at 45°C, or

(iii) 6X SSC at 68°C, 100 mg/ml denatured fish sperm DNA, or

(iv) 6X SSC, 0.5% SDS, 100 mg/ml denatured, fragmented salmon sperm DNA at 68°C, or

(v) 6XSSC, 0.5% SDS, 100 mg/ml denatured, fragmented salmon sperm DNA, 50% formamide at 42°C, or

(vi) 50% formamide, 4X SSC at 42°C, or

(vii) 50% (vol/vol) formamide, 0.1% bovine serum albumin, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 50 mM sodium phosphate buffer pH 6.5, 750 mM NaCl, 75 mM sodium citrate at 42°C, or

(viii) 2X or 4X SSC at 50°C (moderate conditions), or

(ix) 30 to 40% formamide, 2X or 4X SSC at 42° (moderate conditions).

(2) Washing step for 10 minutes each with for example

(i) 0.015 M NaCl/0.0015 M sodium citrate/0.1% SDS at 50°C, or

(ii) 0.1X SSC at 65°C, or

(iii) 0.1X SSC, 0.5% SDS at 68°C, or

(iv) 0.1X SSC, 0.5% SDS, 50% formamide at 42°C, or

(v) 0.2X SSC, 0.1% SDS at 42°C, or

(vi) 2X SSC at 65°C (moderate conditions).

In a preferred embodiment of the process of the invention there is introduction of nucleic acids which encode a ketolase comprising the amino acid sequence SEQ ID NO: 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 50%, preferably at least 60%, preferably at least 65%, preferably at least 70%, more preferably at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, particularly preferably at least 98%, at the amino acid level with the sequence SEQ ID NO: 2.

It is moreover possible for the ketolase sequence to be a natural one which can be found as described above by identity comparison of the sequences from other organisms, or for the ketolase sequence to be an artificial one which has been modified starting from the sequence SEQ ID NO: 2 by artificial variation, for example by substitution, insertion or deletion of amino acids.

The term "substitution" means in the description substitution of one or more amino acids by one or more amino acids. So-called conservative substitutions are preferably carried out, in which

the replaced amino acid has a similar property to the original amino acid, for example substitution of Glu by Asp, Gln by Asn, Val by Ile, Leu by Ile, Ser by Thr.

Deletion is the replacement of an amino acid by a direct linkage. Preferred positions for deletions are the termini of the polypeptide and the linkages between the individual protein domains.

Insertions are introductions of amino acids into the polypeptide chain, with formal replacement of a direct linkage by one or more amino acids.

Identity between two proteins means the identity of the amino acids over the entire length of each protein, in particular the identity calculated by comparison using the vector NTI suite 7.1 software supplied by Informax (USA) using the clustal method (Higgins DG, Sharp PM. Fast and sensitive multiple sequence alignments on a microcomputer. Comput Appl. Biosci. 1989 Apr; 5(2):151-1), setting the following parameters:

Multiple alignment parameter:	
Gap opening penalty	10
Gap extension penalty	10
Gap separation penalty range	8
Gap separation penalty	off
% identity for alignment delay	40
Residue specific gaps	off
Hydrophilic residue gap	off
Transition weighing	0

Pairwise alignment parameter:	
FAST algorithm on	
K-tuple size	1
Gap penalty	3
Window size	5
Number of best diagonals	5

The ketolase having an identity of at least 42% at the amino acid level with the sequence SEQ ID NO: 2 accordingly means a ketolase which, on comparison of its sequence with the sequence SEQ ID NO: 2, in particular using the above program logarithm with the above set of parameters, has an identity of at least 42%.

For example, using the above program logarithm with the above set of parameters, the

sequence of the ketolase from *Nostoc punctiforme ATTC 29133* (SEQ ID NO: 4) shows an identity of 64% with the sequence of the ketolase from *Nostoc sp. Strain PCC7120* (SEQ ID NO: 2).

The sequence of the second ketolase from *Nostoc punctiforme ATCC 29133* (SEQ ID NO: 6) has, for example, an identity of 58% with the sequence of the ketolase from *Nostoc sp. Strain PCC7120* (SEQ ID NO: 2).

The sequence of the ketolase from *Synechococcus sp. WH 8102* (SEQ ID NO: 47) has, for example, an identity of 44% with the sequence of the ketolase from *Nostoc sp. Strain PCC7120* (SEQ ID NO: 2).

Suitable nucleic acid sequences can be obtained for example by back-translation of the polypeptide sequence in accordance with the genetic code.

The codons preferably used for this purpose are those frequently used in accordance with the organism-specific codon usage. The codon usage can easily be found by means of computer analyses of other, known genes in the relevant organisms.

In a particularly preferred embodiment, a nucleic acid comprising the sequence SEQ ID NO: 1 is introduced into the organism.

All the aforementioned ketolase genes can moreover be prepared in a manner known per se by chemical synthesis from the nucleotide units such as, for example, by fragment condensation of individual overlapping, complementary nucleic acid units of the double helix. Chemical synthesis of oligonucleotides is possible, for example, in a known manner by the phosphoramidite method (Voet, Voet, 2nd edition, Wiley Press New York, pages 896-897). Addition of synthetic oligonucleotides and filling in of gaps using the Klenow fragment of DNA polymerase and ligation reactions, and general cloning methods, are described in Sambrook et al. (1989), Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press.

The identity shown by the sequence of the ketolase from *Nostoc sp. Strain PCC7120* (SEQ ID NO: 2) with the sequences of the ketolases used in the prior art processes is 39% (*Agrobacterium aurantiacum* (EP 735 137), Accession No. D58420), 40% (*Alcaligenes sp. PC-1* (EP 735137), Accession No. D58422) and 20 to 21% (*Haematococcus pluvialis Flotow em. Wille* and *Haematoccus pluvialis, NIES 144* (EP 725137, WO 98/18910 and Lotan et al, FEBS Letters 1995, 364, 125 128), Accession No. X86782 and D45881).

In a preferred embodiment, organisms which have an increased hydroxylase activity and/or

β-cyclase activity in addition to the increased ketolase activity compared with the wild type are cultivated.

Hydroxylase activity means the enzymic activity of a hydroxylase.

A hydroxylase means a protein having the enzymatic activity of introducing a hydroxyl group on the, optionally substituted, β-ionone ring of carotenoids.

In particular, a hydroxylase means a protein having the enzymatic activity of converting β-carotene into zeaxanthin or canthaxanthin into astaxanthin.

Accordingly, hydroxylase activity means the amount of β-carotene or canthaxanthin converted, or amount of zeaxanthin or astaxanthin produced, by the hydroxylase protein.

Thus, when the hydroxylase activity is increased compared with the wild type, the amount of β-carotene or canthaxantin converted or the amount of zeaxanthin or astaxanthin produced in a particular time by the hydroxylase protein is increased compared with the wild type.

This increase in the hydroxylase activity is preferably at least 5%, further preferably at least 20%, further preferably at least 50%, further preferably at least 100%, more preferably at least 300%, even more preferably at least 500%, in particular at least 600%, of the hydroxylase activity of the wild type.

β-Cyclase activity means the enzymic activity of a β-cyclase.

A β-cyclase means a protein having the enzymatic activity of converting a terminal linear lycopene residue into a β-ionone ring.

In particular, a β-cyclase means a protein having the enzymatic activity of converting γ-carotene into β-carotene.

Accordingly, a β-cyclase activity means the amount of γ-carotene converted or the amount of β-carotene produced in a particular time by the β-cyclase protein.

Thus, when the β-cyclase activity is increased compared with the wild type, the amount of lycopene or γ-carotene converted or the amount of γ-carotene produced from lycopene or the amount of β-carotene produced from γ-carotene by the β-cyclase protein in a particular time is increased compared with the wild type.

This increase in the β-cyclase activity is preferably at least 5%, further preferably at least 20%, further preferably at least 50%, further preferably at least 100%, more preferably at least 300%, even more preferably at least 500%, in particular at least 600%, of the β-cyclase activity of the wild type.

The hydroxylase activity in the genetically modified organisms of the invention and in wild-type and reference organisms is preferably determined under the following conditions:

The hydroxylase activity is determined by the method of Bouvier et al. (Biochim. Biophys. Acta 1391 (1998), 320-328) *in vitro*. Ferredoxin, ferredoxin-NADP oxidoreductase, catalase, NADPH and β-carotene with mono- and digalactosyl glycerides are added to a defined amount of organism extract.

The hydroxylase activity is particularly preferably determined under the following conditions of Bouvier, Keller, d'Harlingue and Camara (Xanthophyll biosynthesis: molecular and functional characterization of carotenoid hydroxylases from pepper fruits (Capsicum annuum L.); Biochim. Biophys. Acta 1391 (1998), 320-328):

The *in vitro* assay is carried out in a volume of 0.250 ml. The mixture contains 50 mM potassium phosphate (pH 7.6), 0.025 mg of spinach ferredoxin, 0.5 units of spinach ferredoxin-$NADP^+$ oxidoreductase, 0.25 mM NADPH, 0.010 mg of beta-carotene (emulsified in 0.1 mg of Tween 80), 0.05 mM of a mixture of mono- and digalactosyl glycerides (1:1), 1 unit of catalysis, 200 mono- and digalactosyl glycerides (1:1), 0.2 mg of bovine serum albumin and organism extract in a different volume. The reaction mixture is incubated at 30°C for 2 hours. The reaction products are extracted with organic solvents such as acetone or chloroform/methanol (2:1) and determined by HPLC.

The β-cyclase activity in the genetically modified organisms of the invention and in wild-type and reference organisms is preferably determined under the following conditions:

The β-cyclase activity is determined by the method of Fraser and Sandmann (Biochem. Biophys. Res. Comm. 185(1) (1992) 9 15) *in vitro*. Potassium phosphate is added as buffer (pH 7.6), lycopene as substrate, paprika stromal protein, $NADP^+$, NADPH and ATP to a defined amount of organism extract.

The β-cyclase activity is particularly preferably determined under the following conditions of Bouvier, d'Harlingue and Camara (Molecular Analysis of carotenoid cyclase inhibition; Arch.

Biochem. Biophys. 346(1) (1997) 53-64):

The in vitro assay is carried out in a volume of 250 µl volume. The mixture contains 50 mM potassium phosphate (pH 7.6), various amounts of organism extract, 20 nM lycopene, 250 µg of paprika chromoplastid stromal protein, 0.2 mM $NADP^+$, 0.2 mM NADPH and 1 mM ATP. NADP/NADPH and ATP are dissolved in 10 ml of ethanol with 1 mg of Tween 80 immediately before addition to the incubation medium. After a reaction time of 60 minutes at 30°C, the reaction is stopped by adding chloroform/methanol (2:1). The reaction products extracted into chloroform are analyzed by HPLC.

An alternative assay with radioactive substrate is described in Fraser and Sandmann (Biochem. Biophys. Res. Comm. 185(1) (1992) 9-15).

The hydroxylase activity and/or β-cyclase activity can be increased in various ways, for example by switching off inhibitory regulatory mechanisms at the expression and protein level or by increasing the gene expression of nucleic acids encoding a hydroxylase, and/or of nucleic acids encoding a β-cyclase, compared with the wild type.

The gene expression of nucleic acids encoding a hydroxylase, and/or the gene expression of the nucleic acid encoding a β-cyclase, compared with the wild type, can likewise be increased in various ways, for example by inducing the hydroxylase gene and/or β-cyclase gene by activators or by introducing one or more hydroxylase gene copies and/or β-cyclase gene copies, i.e. by introducing at least one nucleic acid encoding a hydroxylase, and/or at least one nucleic acid encoding a β-cyclase, into the organism.

Increasing the gene expression of a nucleic acid encoding a hydroxylase and/or β-cyclase also means according to the invention manipulation of the expression of the organism's own endogenous hydroxylase and/or β-cyclase.

This can be achieved for example by modifying the promoter DNA sequence for genes encoding hydroxylases and/or β-cyclases. Such a modification, resulting in an increased expression rate of the gene, can be effected for example by deletion or insertion of DNA sequences.

It is possible, as described above, to modify the expression of the endogenous hydroxylase and/or β-cyclase by application of exogenous stimuli. This can be effected by particular physiological conditions, i.e. by application of foreign substances.

A further possibility for achieving a modified or increased expression of an endogenous

hydroxylase and/or β-cyclase gene is through interaction of a regulator protein which does not occur in the untransformed organism with the promoter of this gene.

Such a regulator may be a chimeric protein consisting of a DNA-binding domain and of a transcription activator domain as described, for example, in WO 96/06166.

In a preferred embodiment, the gene expression of a nucleic acid encoding a hydroxylase, and/or the gene expression of a nucleic acid encoding a β-cyclase, is increased by introducing at least one nucleic acid encoding a hydroxylase, and/or by introducing at least one nucleic acid encoding a β-cyclase, into the organism.

It is possible to use for this purpose in principle any hydroxylase gene or any β-cyclase gene, i.e. any nucleic acid which encodes a hydroxylase and any nucleic acid which encodes a β-cyclase.

In the case of genomic hydroxylase or β-cyclase nucleic acid sequences from eukaryotic sources which comprise introns, it is preferred to use nucleic acid sequences which have already been processed, such as the corresponding cDNAs, in the case where the host organism is unable or cannot be made able to express the corresponding hydroxylase or β-cyclase.

One example of a hydroxylase gene is a nucleic acid encoding a hydroxylase from *Haematococcus pluvialis* (Accession AX038729, WO 0061764); (nucleic acid: SEQ ID NO: 15, protein: SEQ ID NO: 16).

One example of a β-cyclase gene is a nucleic acid encoding a β-cyclase from tomato (Accession X86452) (nucleic acid: SEQ ID NO: 17, protein: SEQ ID NO: 18).

Thus, in this preferred embodiment, at least one further hydroxylase gene and/or β-cyclase gene is present in the preferred transgenic organisms of the invention compared with the wild type.

In this preferred embodiment, the genetically modified organism has for example at least one exogenous nucleic acid encoding a hydroxylase, or at least two endogenous nucleic acids encoding a hydroxylase and/or at least one exogenous nucleic acid encoding a β-cyclase, or at least two endogenous nucleic acids encoding a β-cyclase.

The hydroxylase genes preferably used in the preferred embodiment described above are nucleic acids encoding proteins comprising the amino acid sequence SEQ ID NO: 16 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino

acids and which have an identity of at least 30%, preferably at least 50%, more preferably at least 70%, even more preferably at least 90%, most preferably at least 95%, at the amino acid level with the sequence SEQ ID NO: 16, and which have the enzymatic property of a hydroxylase.

Further examples of hydroxylases and hydroxylase genes can be easily found for example from various organisms whose genomic sequence is known as described above by homology comparisons of the amino acid sequences or of the corresponding back-translated nucleic acid sequences from databases with SEQ ID. NO: 16.

Further examples of hydroxylases and hydroxylase genes can easily be found in a manner known per se in addition for example starting from the sequence SEQ ID NO: 15 from various organisms whose genomic sequence is unknown, as described above, by hybridization and PCR techniques.

In a further particularly preferred embodiment, nucleic acids which encode proteins comprising the amino acid sequence of the hydroxylase of the sequence SEQ ID NO: 16 are introduced into organisms to increase the hydroxylase activity.

Suitable nucleic acid sequences can be obtained for example by back-translation of the polypeptide sequence in accordance with the genetic code.

The codons used for this purpose are preferably those frequently used in accordance with the organism-specific codon usage. This codon usage can easily be found by means of computer analyses of other, known genes of the relevant organisms.

In a particularly preferred embodiment, a nucleic acid comprising the sequence SEQ. ID. NO: 15 is introduced into the organism.

The β-cyclase genes preferably used in the preferred embodiment described above are nucleic acids which encode proteins comprising the amino acid sequence SEQ ID NO: 18 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 30%, preferably at least 50%, more preferably at least 70%, even more preferably at least 90%, most preferably at least 95%, at the amino acid level with the sequence SEQ ID NO: 18, and which has the enzymatic property of a β-cyclase.

Further examples of β-cyclases and β-cyclase genes can easily be found for example from various organisms whose genomic sequence is known as described above by homology comparisons of the amino acid sequences or of the corresponding back-translated nucleic acid

sequences from databases with the SEQ ID NO: 18.

Further examples of β-cyclases and β-cyclase genes can easily be found in a manner known per se in addition for example starting from the sequence SEQ ID NO: 17 from various organisms whose genomic sequence is unknown by hybridization and PCR techniques.

In a further particularly preferred embodiment, nucleic acids which encode proteins comprising the amino acid sequence of β-cyclase of the sequence SEQ. ID. NO: 18 are introduced into organisms to increase the β-cyclase activity.

Suitable nucleic acid sequences can be obtained for example by back-translation of the polypeptide sequence in accordance with the genetic code.

The codons preferably used for this purpose are those frequently used in accordance with the organ-specific codon usage. This codon usage can easily be found by means of computer analyses of other, known genes of the relevant organisms.

In a particularly preferred embodiment, a nucleic acid comprising the sequence SEQ. ID. NO: 17 is introduced into the organism.

All the aforementioned hydroxylase genes or β-cyclase genes can moreover be prepared in a manner known per se by chemical synthesis from the nucleotide units such as, for example, by fragment condensation of individual overlapping, complementary nucleic acid units of the double helix. Chemical synthesis of oligonucleotides is possible, for example, in a known manner by the phosphoramidite method (Voet, Voet, 2nd edition, Wiley Press New York, pages 896-897). Addition of synthetic oligonucleotides and filling in of gaps using the Klenow fragment of DNA polymerase and ligation reactions, and general cloning methods, are described in Sambrook et al. (1989), Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press.

The genetically modified organisms particularly preferably used in the process of the invention have the following combinations of genetic modifications:

genetically modified organisms which have, compared with the wild type, an increased or caused ketolase activity and an increased hydroxylase activity,

genetically modified organisms which have, compared with the wild type, an increased or caused ketolase activity and an increased β-cyclase activity and

genetically modified organisms which have, compared with the wild type, an increased or caused ketolase activity and an increased hydroxylase activity and an increased β-cyclase activity.

These genetically modified organisms can be produced as described hereinafter for example by introducing individual nucleic acid constructs (expression cassettes) or by introducing multiple constructs which comprise up to two or three of the described activities.

Organisms preferably mean according to the invention organisms which are able as wild-type or starting organisms naturally or through genetic complementation and/or reregulation of metabolic pathways to produce carotenoids, in particular β-carotene and/or zeaxanthin and/or neoxanthin and/or violaxanthin and/or lutein.

Further preferred organisms already have as wild-type or starting organisms a hydroxylase activity and are thus able as wild-type or starting organisms to produce zeaxanthin.

Preferred organisms are plants or microorganisms such as, for example, bacteria, yeasts, algae or fungi.

Bacteria which can be used are both bacteria which are able, because of the introduction of genes of carotenoid biosynthesis of a carotenoid-producing organism, to synthesize xanthophylls, such as, for example, bacteria of the genus *Escherichia*, which comprise for example crt genes from *Erwinia*, and bacteria which are intrinsically able to synthesize xanthophylls, such as, for example, bacteria of the genus *Erwinia*, *Agrobacterium*, *Flavobacterium*, *Alcaligenes*, *Paracoccus*, *Nostoc* or cyanobacteria of the genus *Synechocystis*.

Preferred bacteria are *Escherichia coli, Erwinia herbicola*, *Erwinia uredovora, Agrobacterium aurantiacum, Alcaligenes sp. PC-1, Flavobacterium* sp. strain R1534, the cyanobacterium *Synechocystis* sp. PCC6803, *Paracoccus marcusii* or *Paracoccus carotinifaciens*.

Preferred yeasts are *Candida*, *Saccharomyces*, *Hansenula*, *Pichia* or *Phaffia*. Particularly preferred yeasts are *Xanthophyllomyces dendrorhous* or *Phaffia rhodozyma*.

Preferred fungi are *Aspergillus, Trichoderma, Ashbya, Neurospora, Blakeslea, Phycomyces, Fusarium* or other fungi described in Indian Chem. Engr. Section B. Vol. 37, No. 1, 2 (1995) on page 15, table 6.

Preferred algae are green algae such as, for example, algae of the genus *Haematococcus*, *Phaedactylum tricornatum, Volvox* or *Dunaliella*. Particularly preferred algae are *Haematococcus*

pluvialis or *Dunaliella bardawil.*

Further microorganisms which can be used and the production thereof for carrying out the process of the invention are disclosed for example in DE-A-199 16 140, which is incorporated herein by reference.

Particularly preferred plants are plants selected from the families Ranunculaceae, Berberidaceae, Papaveraceae, Cannabaceae, Rosaceae, Fabaceae, Linaceae, Vitaceae, Brassicaceae, Cucurbitaceae, Primulaceae, Caryophyllaceae, Amaranthaceae, Gentianaceae, Geraniaceae, Caprifoliaceae, Oleaceae, Tropaeolaceae, Solanaceae, Scrophulariaceae, Asteraceae, Liliaceae, Amaryllidaceae, Poaceae, Orchidaceae, Malvaceae, Illiaceae or Lamiaceae.

Very particularly preferred plants are selected from the group of plant genera *Marigold, Tagetes erecta, Tagetes patula, Acacia, Aconitum, Adonis, Arnica, Aquilegia, Aster, Astragalus, Bignonia, Calendula, Caltha, Campanula, Canna, Centaurea, Cheiranthus, Chrysanthemum, Citrus, Crepis, Crocus, Curcurbita, Cytisus, Delonia, Delphinium, Dianthus, Dimorphotheca, Doronicum, Eschscholtzia, Forsythia, Fremontia, Gazania, Gelsemium, Genista, Gentiana, Geranium, Gerbera, Geum, Grevillea, Helenium, Helianthus, Hepatica, Heracleum, Hibiscus, Heliopsis, Hypericum, Hypochoeris, Impatiens, Iris, Jacaranda, Kerria, Laburnum, Lathyrus, Leontodon, Lilium, Linum, Lotus, Lycopersicon, Lysimachia, Maratia, Medicago, Mimulus, Narcissus, Oenothera, Osmanthus, Petunia, Photinia, Physalis, Phyteuma, Potentilla, Pyracantha, Ranunculus, Rhododendron, Rosa, Rudbeckia, Senecio, Silene, Silphium, Sinapsis, Sorbus, Spartium, Tecoma, Torenia, Tragopogon, Trollius, Tropaeolum, Tulipa, Tussilago, Ulex, Viola* or *Zinnia,* particularly preferably selected from the group of plant genera Marigold, Tagetes erecta, Tagetes patula, Lycopersicon, Rosa, Calendula, Physalis, Medicago, Helianthus, Chrysanthemum, Aster, Tulipa, Narcissus, Petunia, Geranium, Tropaeolum or Adonis.

In the process of the invention for preparing ketocarotenoids, the step of cultivating the genetically modified organisms is preferably followed by a harvesting of the organisms and further preferably in addition by an isolation of ketocarotenoids from the organisms.

The harvesting of the organisms takes place in a manner known per se appropriate for the particular organism. Microorganisms such as bacteria, yeasts, algae or fungi or plant cells cultivated by fermentation in liquid nutrient media can be removed for example by centrifugation, decantation or filtration. Plants are grown on nutrient media and appropriately harvested in a manner known per se.

The genetically modified microorganisms are preferably cultivated in the presence of oxygen at a

cultivation temperature of at least about 20°C, such as for example, 20°C to 40°C, and at a pH of about 6 to 9. In the case of genetically modified microorganisms, the microorganisms are preferably initially cultivated in the presence of oxygen and in a complex medium such as, for example, TB or LB medium at a cultivation temperature of about 20°C or more, and at a pH of about 6 to 9, until a sufficient cell density is reached. In order to be able to control the oxidation reaction better, it is preferred to use an inducible promoter. The cultivation is continued after induction of ketolase expression in the presence of oxygen for example for 12 hours to 3 days.

The ketocarotenoids are isolated from the harvested biomass in a manner known per se, for example by extraction and, where appropriate, further chemical or physical purification processes such as, for example, precipitation methods, crystallography, thermal separation processes, such as rectification processes or physical separation processes such as, for example, chromatography.

As mentioned below, the ketocarotenoids can be specifically produced in the genetically modified plants of the invention preferably in various plant tissues such as, for example, seeds, leaves, fruits, flowers, especially in petals.

Ketocarotenoids are isolated from the harvested petals in a manner known per se, for example by drying and subsequent extraction and, where appropriate, further chemical or physical purification processes such as, for example, precipitation methods, crystallography, thermal separation processes such as rectification processes or physical separation processes such as, for example, chromatography. Ketocarotenoids are isolated from petals for example preferably by organic solvents such as acetone, hexane, ether or tert-methyl butyl ether.

Further processes for isolating ketocarotenoids, especially from petals, are described for example in Egger and Kleinig (Phytochemistry (1967) 6, 437-440) and Egger (Phytochemistry (1965) 4, 609-618).

The ketocarotenoids are preferably selected from the group of astaxanthin, canthaxanthin, echinenone, 3-hydroxyechinenone, 3'-hydroxyechinenone, adonirubin and adonixanthin.

Astaxanthin is a particularly preferred ketocarotenoid.

Depending on the organism used, the ketocarotenoids are obtained in free form or as fatty acid ester.

In plant petals, the ketocarotenoids are obtained in the process of the invention in the form of their mono- or diesters with fatty acids. Some examples of detected fatty acids are myristic acid,

palmitic acid, stearic acid, oleic acid, linolenic acid and lauric acid (Kamata and Simpson (1987) Comp. Biochem. Physiol. Vol. 86B(3), 587-591).

The ketocarotenoids can be produced in the whole plant or, in a preferred embodiment, specifically in plant tissues containing chromoplasts. Examples of preferred plant tissues are roots, seeds, leaves, fruits, flowers and, in particular nectaries and petals.

In a particularly preferred embodiment of the process of the invention, genetically modified plants which show the highest rate of expression of a ketolase in flowers are used.

This is preferably achieved through the ketolase gene expression being under the control of a flower-specific promoter. For this purpose, for example, the nucleic acids described above are introduced into the plant, as described in detail below, in a nucleic acid construct functionally linked to a flower-specific promoter.

In a further, particularly preferred embodiment of the process of the invention, genetically modified plants which show the highest rate of expression of a ketolase in fruits are used.

This is preferably achieved through the ketolase gene expression being under the control of a fruit-specific promoter. For this purpose, for example, the nucleic acids described above are introduced into the plant, as described in detail below, in a nucleic acid construct functionally linked to a fruit-specific promoter.

In a further, particularly preferred embodiment of the process of the invention, genetically modified seeds which show the highest rate of expression of a ketolase in seeds are used.

This is preferably achieved through the ketolase gene expression being under the control of a seed-specific promoter. For this purpose, for example, the nucleic acids described above are introduced into the plant, as described in detail below, in a nucleic acid construct functionally linked to a seed-specific promoter.

The targeting into the chromoplasts is effected by a functionally linked plastid transit peptide.

The production of genetically modified plants with increased or caused ketolase activity is described by way of example below. Further activities such as, for example, the hydroxylase activity and/or the β-cyclase activity can be increased analogously using nucleic acid sequences encoding a hydroxylase or β-cyclase in place of nucleic acid sequences encoding a ketolase. The transformation can be effected in the combinations of genetic modifications singly or by multiple constructs.

The transgenic plants are preferably produced by transformation of the starting plants using a nucleic acid construct which comprises the nucleic acids described above encoding a ketolase, which are functionally linked to one or more regulatory signals which ensure transcription and translation in plants.

These nucleic acid constructs in which the coding nucleic acid sequence is functionally linked to one or more regulatory signals which ensure transcription and translation in plants are also called expression cassettes below.

The regulatory signals preferably comprise one or more promoters which ensure transcription and translation in plants.

The expression cassettes comprise regulatory signals, i.e. regulating nucleic acid sequences which control the expression of the coding sequence in the host cell. In a preferred embodiment, an expression cassette comprises a promoter upstream, i.e. at the 5' end of the coding sequence, and a polyadenylation signal downstream, i.e. at the 3' end, and, where appropriate, further regulatory elements which are operatively linked to the coding sequence, located in between, for at least one of the genes described above. Operative linkage means the sequential arrangement of promoter, coding sequence, terminator and, where appropriate, further regulatory elements in such a way that each of the regulatory elements is able to carry out its function as intended in the expression of the coding sequence.

The preferred nucleic acid constructs, expression cassettes and vectors for plants and processes for producing transgenic plants, and the transgenic plants themselves, are described by way of example below.

The sequences which are preferred for the operative linkage, but are not restricted thereto, are targeting sequences to ensure the subcellular localization in the apoplast, in the vacuole, in plastids, in the mitochondrion, in the endoplasmic reticulum (ER), in the cell nucleus, in elaioplasts or other compartments and translation enhancers such as the 5' leader sequence from tobacco mosaic virus (Gallie et al., Nucl. Acids Res. 15 (1987), 8693 -8711).

A suitable promoter for the expression cassette is in principle any promoter able to control the expression of foreign genes in plants.

"Constitutive" promoter means promoters which ensure expression in numerous, preferably all, tissues over a relatively wide period during development of the plant, preferably at all times during development of the plant.

Preferably used is, in particular, a plant promoter or a promoter derived from a plant virus. Particular preference is given to the CaMV promoter of the 35S transcript of cauliflower mosaic virus (Franck et al. (1980) Cell 21:285-294; Odell et al. (1985) Nature 313:810-812; Shewmaker et al. (1985) Virology 140:281-288; Gardner et al. (1986) Plant Mol Biol 6:221-228), the 19S CaMV promoter (US 5,352,605; WO 84/02913; Benfey et al. (1989) EMBO J 8:2195-2202), the triose phosphate translocator (TPT) promoter from *Arabidopsis thaliana* Acc. No. AB006698, base pair 53242 to 55281; the gene starting at bp 55282 is anotated as "phosphate/triose phosphate translocator", or the 34S promoter from figwort mosaic virus Acc. No. X16673, base pair 1 to 554.

A further suitable constitutive promoter is the pds promoter (Pecker et al. (1992) Proc. Natl. Acad. Sci USA 89: 4962-4966) or the rubisco small subunit (SSU) promoter (US 4,962,028), the legumin B promoter (GenBank Acc. No. X03677), the agrobacterium nopaline synthase promoter, the TR dual promoter, the agrobacterium OCS (octopine synthase) promoter, the ubiquitin promoter (Holtorf S et al. (1995) Plant Mol Biol 29:637-639), the ubiquitin 1 promoter (Christensen et al. (1992) Plant Mol Biol 18:675-689; Bruce et al. (1989) Proc Natl Acad Sci USA 86:9692-9696), the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (US 5,683,439), the promoters of the vacuolar ATPase subunits or the promoter of a proline-rich protein from wheat (WO 91/13991), the Pnit promoter (Y07648.L, Hillebrand et al. (1998), Plant. Mol. Biol. 36, 89-99, Hillebrand et al. (1996), Gene, 170, 197-200) and further promoters of genes whose constitutive expression in plants is known to the skilled worker.

The expression cassettes may also comprise a chemically inducible promoter (review article: Gatz et al. (1997) Annu Rev Plant Physiol Plant Mol Biol 48:89-108) by which expression of the ketolase gene in the plant can be controlled at a particular time. Promoters of this type, such as, for example, the PRP1 promoter (Ward et al. (1993) Plant Mol Biol 22:361-366), a salicylic acid-inducible promoter (WO 95/19443), a benzenesulfonamide-inducible promoter (EP 0 388 186), a tetracycline-inducible promoter (Gatz et al. (1992) Plant J 2:397-404), an abscisic acid-inducible promoter (EP 0 335 528) or an ethanol- or cyclohexanone-inducible promoter (WO 93/21334), can likewise be used.

Promoters which are further preferred are those induced by biotic or abiotic stress, such as, for example, the pathogen-inducible promoter of the PRP1 gene (Ward et al. (1993) Plant Mol Biol 22:361-366), the heat-inducible tomato hsp70 or hsp80 promoter (US 5,187,267), the cold-inducible potato alpha-amylase promoter (WO 96/12814), the light-inducible PPDK promoter or the wound-induced pinII promoter (EP375091).

Pathogen-inducible promoters include those of genes which are induced as a result of pathogen

attack, such as, for example, genes of PR proteins, SAR proteins, β-1,3-glucanase, chitinase etc. (for example Redolfi et al. (1983) Neth J Plant Pathol 89:245-254; Uknes, et al. (1992) The Plant Cell 4:645-656; Van Loon (1985) Plant Mol Viral 4:111-116; Marineau et al. (1987) Plant Mol Biol 9:335-342; Matton et al. (1987) Molecular Plant-Microbe Interactions 2:325-342; Somssich et al. (1986) Proc Natl Acad Sci USA 83:2427-2430; Somssich et al. (1988) Mol Gen Genetics 2:93-98; Chen et al. (1996) Plant J 10:955-966; Zhang and Sing (1994) Proc Natl Acad Sci USA 91:2507-2511; Warner, et al. (1993) Plant J 3:191-201; Siebertz et al. (1989) Plant Cell 1:961-968 (1989).

Also included as wound-inducible promoters such as that of the pinII gene (Ryan (1990) Ann Rev Phytopath 28:425-449; Duan et al. (1996) Nat Biotech 14:494-498), of the wun1 and wun2 genes (US 5,428,148), of the win1 and win2 genes (Stanford et al. (1989) Mol Gen Genet 215:200-208), of the systemin gene (McGurl et al. (1992) Science 225:1570-1573), of the WIP1 gene (Rohmeier et al. (1993) Plant Mol Biol 22:783-792; Ekelkamp et al. (1993) FEBS Letters 323:73-76), of the MPI gene (Corderok et al. (1994) The Plant J 6(2):141-150) and the like.

Examples of further suitable promoters are fruit ripening-specific promoters such as, for example, the tomato fruit ripening-specific promoter (WO 94/21794, EP 409 625). Development-dependent promoters include some of the tissue-specific promoters because the formation of some tissues naturally depends on development.

Further particularly preferred promoters are those which ensure expression in tissues or parts of plant in which, for example, the biosynthesis of ketocarotenoids or precursors thereof takes place. Preferred examples are promoters having specificities for anthers, ovaries, petals, sepals, flowers, leaves, stalks, seeds and roots and combinations thereof.

Examples of promoters specific for tubers, storage roots or roots are the patatin promoter class I (B33) or the potato cathepsin D inhibitor promoter.

Examples of leaf-specific promoters are the promoter of the potato cytosolic FBPase (WO 97/05900), the rubisco (ribulose-1,5-bisphosphate carboxylase) SSU promoter (small subunit) or the potato ST-LSI promoter (Stockhaus et al. (1989) EMBO J 8:2445-2451).

Examples of flower-specific promoters are the phytoene synthase promoter (WO 92/16635) or the promoter of the P-rr gene (WO 98/22593), the Arabidopsis thaliana AP3 promoter (see example 5), the CHRC promoter (chromoplast-specific carotenoid-associated protein (CHRC) gene promoter from Cucumis sativus Acc. No. AF099501, base pair 1 to 1532), the EPSP synthase promoter (5-enolpyruvylshikimate-3-phosphate synthase gene promoter from Petunia hybrida, Acc. No. M37029, base pair 1 to 1788), the PDS promoter (phytoene desaturase gene

promoter from Solanum lycopersicum, Acc. No. U46919, base pair 1 to 2078), the DFR-A promoter (dihydroflavonol 4-reductase gene A promoter from Petunia hybrida, Acc. No. X79723, base pair 32 to 1902) or the FBP1 promoter (floral binding protein 1 gene promoter from Petunia hybrida, Acc. No. L10115, base pair 52 to 1069).

Examples of anther-specific promoters are the 5126 promoter (US 5,689,049, US 5,689,051), the glob-1 promoter or the g-zein promoter.

Examples of seed-specific promoters are the ACP05 promoter (acyl carrier protein gene, WO 9218634), the *Arabidopsis* AtS1 and AtS3 promoters (WO 9920775), the *Vicia faba* LeB4 promoter (WO 9729200 and US 06403371), the *Brassica napus* napin promoter (US 5608152; EP 255378; US 5420034), the *Vicia faba* SBP promoter (DE 9903432) or the maize End1 and End2 promoters (WO 0011177).

Further promoters suitable for expression in plants are described in Rogers et al. (1987) Meth in Enzymol 153:253-277; Schardl et al. (1987) Gene 61:1-11 and Berger et al. (1989) Proc Natl Acad Sci USA 86:8402-8406.

Particularly preferred in the process of the invention are constitutive, seed-specific, fruit-specific, flower-specific and, in particular, petal-specific promoters.

The present invention therefore relates in particular to a nucleic acid construct comprising functionally linked a flower-specific or, in particular, a petal-specific promoter and a nucleic acid encoding a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2.

An expression cassette is preferably produced by fusing a suitable promoter to a nucleic acid, described above, encoding a ketolase, and preferably to a nucleic acid which is inserted between promoter and nucleic acid sequence and which codes for a plastid-specific transit peptide, and to a polyadenylation signal by conventional recombination and cloning techniques as described, for example in T. Maniatis, E.F. Fritsch and J. Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989) and in T.J. Silhavy, M.L. Berman and L.W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and in Ausubel, F.M. et al., Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley-Interscience (1987).

The preferably inserted nucleic acids encoding a plastid transit peptide ensure localization in plastids and, in particular, in chromoplasts.

It is also possible to use expression cassettes whose nucleic acid sequence codes for a ketolase fusion protein, where part of the fusion protein is a transit peptide which controls the translocation of the polypeptide. Transit peptides which are specific for chromoplasts and which are eliminated enzymatically from the ketolase part after translocation of the ketolase into the chromoplasts.

The particularly preferred transit peptide is derived from the *Nicotiana tabacum* plastid transketolase or another transit peptide (e.g. the transit peptide of the small subunit of rubisco (rbcS) or of the ferredoxin NADP oxidoreductase, as well as the isopentenyl-pyrophosphate isomerase 2) or its functional equivalent.

Particular preference is given to nucleic acid sequences of three cassettes of the plastid transit peptide of the tobacco plastic transketolase in three reading frames as KpnI/BamHI fragments with an ATG codon in the NcoI cleavage site:

pTP09

KpnI_GGTACCATGGCGTCTTCTTCTTCTCTCACTCTCTCTCAAGCTATCCTCTCTCGTTCTG
TCCCTCGCCATGGCTCTGCCTCTTCTTCTCAACTTTCCCCTTCTTCTCTCACTTTTTCCGGC
CTTAAATCCAATCCCAATATCACCACCTCCCGCCGCCGTACTCCTTCCTCCGCCGCCGCC
GCCGCCGTCGTAAGGTCACCGGCGATTCGTGCCTCAGCTGCAACCGAAACCATAGAGAAA
ACTGAGACTGCGGGATCC_BamHI

pTP10

KpnI_GGTACCATGGCGTCTTCTTCTTCTCTCACTCTCTCTCAAGCTATCCTCTCTCGTTCTG
TCCCTCGCCATGGCTCTGCCTCTTCTTCTCAACTTTCCCCTTCTTCTCTCACTTTTTCCGGC
CTTAAATCCAATCCCAATATCACCACCTCCCGCCGCCGTACTCCTTCCTCCGCCGCCGCC
GCCGCCGTCGTAAGGTCACCGGCGATTCGTGCCTCAGCTGCAACCGAAACCATAGAGAAA
ACTGAGACTGCGCTGGATCC_BamHI

pTP11

KpnI_GGTACCATGGCGTCTTCTTCTTCTCTCACTCTCTCTCAAGCTATCCTCTCTCGTTCTG
TCCCTCGCCATGGCTCTGCCTCTTCTTCTCAACTTTCCCCTTCTTCTCTCACTTTTTCCGGC
CTTAAATCCAATCCCAATATCACCACCTCCCGCCGCCGTACTCCTTCCTCCGCCGCCGCC
GCCGCCGTCGTAAGGTCACCGGCGATTCGTGCCTCAGCTGCAACCGAAACCATAGAGAAA
ACTGAGACTGCGGGGATCC_BamHI

Further examples of a plastid transit peptide are the transit peptide of the Arabidopsis thaliana plastid isopentenyl-pyrophosphate isomerase 2 (IPP-2) and the transit peptide of the small subunit of ribulose-bisphosphate carboxylase (rbcS) from pea (Guerineau, F, Woolston, S, Brooks, L, Mullineaux, P (1988) An expression cassette for targeting foreign proteins into the chloroplasts. Nucl. Acids Res. 16: 11380).

The nucleic acids of the invention can be prepared synthetically or obtained naturally or comprise a mixture of synthetic and natural nucleic acid constituents, and consist of various heterologous gene sections from different organisms.

Preference is given, as described above, to synthetic nucleotide sequences with codons preferred by plants. These codons preferred by plants can be identified from codons with the highest protein frequency which are expressed in most plant species of interest.

For preparing an expression cassette it is possible to manipulate various DNA fragments in order to obtain a nucleotide sequence which expediently reads in the correct direction and is equipped with a correct reading frame. Adaptors or linkers can be attached to the fragments for connecting the DNA fragments to one another.

It is possible and expedient for the promoter and terminator regions to be provided in the direction of transcription with a linker or polylinker which contains one or more restriction sites for inserting this sequence. As a rule, the linker has 1 to 10, usually 1 to 8, preferably 2 to 6, restriction sites. The linker generally has a size of less than 100 bp, frequently less than 60 bp, but at least 5 bp, inside the regulatory regions. The promoter may be both native or homologous and foreign or heterologous to the host plant. The expression cassette preferably comprises in the 5'-3' direction of transcription the promoter, a coding nucleic acid sequence or a nucleic acid construct and a region for termination of transcription. Various termination regions are interchangeable as desired.

Examples of a terminator are the 35S terminator (Guerineau et al. (1988) Nucl Acids Res. 16: 11380), the nos terminator (Depicker A, Stachel S, Dhaese P, Zambryski P, Goodman HM. Nopaline synthase: transcript mapping and DNA sequence. J Mol Appl Genet. 1982;1(6):561-73) or the ocs terminator (Gielen, J, de Beuckeleer, M, Seurinck, J, Debroek, H, de Greve, H, Lemmers, M, van Montagu, M, Schell, J (1984) The complete sequence of the TL-DNA of the Agrobacterium tumefaciens plasmid pTiAch5. EMBO J. 3: 835-846).

It is furthermore possible to employ manipulations which provide appropriate restriction cleavage sites or delete the redundant DNA or restriction cleavage sites. It is possible in relation to

insertions, deletions or substitutions, such as, for example, transitions and transversions, to use *in vitro* mutagenesis, primer repair, restriction or ligation.

It is possible with suitable manipulations, such as, for example, restriction, chewing back or filling in of overhangs for blunt ends, to provide complementary ends of the fragments for ligation.

Preferred polyadenylation signals are plant polyadenylation signals, preferably those which essentially correspond to T-DNA polyadenylation signals from Agrobacterium tumefaciens, especially of gene 3 of the T-DNA (octopine synthase) of the Ti plasmid pTiACH5 (Gielen et al., EMBO J. 3 (1984), 835 ff) or functional equivalents.

The transfer of foreign genes into the genome of a plant is referred to as transformation.

It is possible to use for this purpose methods known per se for the transformation and regeneration of plants from plant tissues or plant cells for transient or stable transformation.

Suitable methods for transforming plants are protoplast transformation by polyethylene glycol-induced DNA uptake, the biolistic method using the gene gun - called the particle bombardment method - electroporation, incubation of dry embryos in DNA-containing solution, microinjection and gene transfer mediated by *Agrobacterium* described above. Said processes are described, for example, in B. Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, edited by S.D. Kung and R. Wu, Academic Press (1993), 128-143 and in Potrykus, Annu. Rev. Plant Physiol. Plant Molec. Biol. 42 (1991), 205-225.

The construct to be expressed is preferably cloned into a vector which is suitable for transforming *Agrobacterium tumefaciens*, for example pBin19 (Bevan et al., Nucl. Acids Res. 12 (1984), 8711) or particularly preferably, pSUN2, pSUN3, pSUN4 or pSUN5 (WO 02/00900).

Agrobacteria transformed with an expression plasmid can be used in a known manner for transforming plants, e.g. bathing wounded leaves or pieces of leaf in a solution of agrobacteria and subsequently cultivating in suitable media.

For the preferred production of genetically modified plants, also referred to as transgenic plants hereinafter, the fused expression cassette which expresses a ketolase is cloned into a vector, for example pBin19 or, in particular, pSUN5 and pSUN3, which is suitable for being transformed into *Agrobacterium tumefaciens*. Agrobacteria transformed with such a vector can then be used in a known manner for transforming plants, in particular crop plants, by bathing wounded leaves

or pieces of leaf in a solution of agrobacteria and subsequently cultivating in suitable media.

The transformation of plants by agrobacteria is disclosed inter alia in F.F. White, Vectors for Gene Transfer in Higher Plants; in Transgenic Plants, Vol. 1, Engineering and Utilization, edited by S.D. Kung and R. Wu, Academic Press, 1993, pages 15-38. Transgenic plants which comprise a gene, integrated into the expression cassette for expression of a nucleic acid encoding a ketolase can be regenerated in a known manner from the transformed cells of the wounded leaves or pieces of leaf.

To transform a host cell with a nucleic acid coding for a ketolase, an expression cassette is incorporated and inserted into a recombinant vector whose vector DNA comprises additional functional regulatory signals, for example sequences for replication or integration. Suitable vectors are described inter alia in "Methods in Plant Molecular Biology and Biotechnology" (CRC Press), chapter 6/7, pages 71-119 (1993).

Using the recombination and cloning techniques quoted above, the expression cassettes can be cloned into suitable vectors which make replication thereof possible for example in *E. coli*. Suitable cloning vectors are, inter alia, pJIT117 (Guerineau et al. (1988) Nucl. Acids Res.16 :11380), pBR332, pUC series, M13mp series and pACYC184. Binary vectors which are able to replicate both in *E. coli* and in agrobacteria are particularly suitable.

The production of the genetically modified microorganisms of the invention is described in more detail below:

The nucleic acids described above, encoding a ketolase or β-hydroxylase or β-cyclase, are preferably incorporated into expression constructs comprising, under the genetic control of regulatory nucleic acid sequences, a nucleic acid sequence coding for an enzyme of the invention; and vectors comprising at least one of these expression constructs.

Such constructs of the invention preferably include a promoter upstream, i.e. at the 5' end of the particular coding sequence, and a terminator sequence downstream, i.e. at the 3' end, and, where appropriate, further customary regulatory elements which are in each case operatively linked to the coding sequence. Operative linkage means the sequential arrangement of promoter, coding sequence, terminator and, where appropriate, further regulatory elements in such a way that each of the regulatory elements is able to carry out its function as intended in the expression of the coding sequence.

Examples of operatively linkable sequences are targeting sequences and translation enhancers, enhancers, polyadenylation signals and the like. Further regulatory elements include selectable

markers, amplification signals, origins of replication and the like.

In addition to the artificial regulatory sequences it is possible for the natural regulatory sequence still to be present in front of the actual structural gene. This natural regulation can be switched off where appropriate, and the expression of the genes increased or reduced, by genetic modification. The gene construct may, however, also have a simpler structure, that is to say no additional regulatory signals are inserted in front of the structural gene, and the natural promoter with its regulation is not deleted. Instead, the natural regulatory sequence is mutated so that regulation no longer takes place, and gene expression is increased or reduced. The nucleic acid sequences may be present in one or more copies in the gene construct.

Examples of promoters which can be used are: cos, tac, trp, tet, trp-tet, lpp, lac, lpp-lac, laclq, T7, T5, T3, gal, trc, ara, SP6, lambda-PR or lambda-PL promoter, which are advantageously used in Gram-negative bacteria; and the Gram-positive promoters amy and SPO2 or the yeast promoters ADC1, MFα, AC, P-60, CYC1, GAPDH. The use of inducible promoters is particularly preferred, such as, for example, light- and, in particular, temperature-inducible promoters such as the P_rP_l promoter.

It is possible in principle for all natural promoters with their regulatory sequences to be used. In addition, it is also possible advantageously to use synthetic promoters.

Said regulatory sequences are intended to make specific expression of the nucleic acid sequences and protein expression possible. This may mean, for example, depending on the host organism, that the gene is expressed or overexpressed only after induction or that it is immediately expressed and/or overexpressed.

The regulatory sequences or factors may moreover preferably influence positively, and thus increase or reduce, expression. Thus, enhancement of the regulatory elements can take place advantageously at the level of transcription by using strong transcription signals such as promoters and/or enhancers. However, it is also possible to enhance translation by, for example, improving the stability of the mRNA.

An expression cassette is produced by fusing a suitable promoter to the above described nucleic acid sequence which encodes a ketolase, β-hydroxylase or β-cyclase and to a terminator signal or polyadenylation signal. Conventional techniques of recombination and cloning are used for this purpose, as described, for example, in T. Maniatis, E.F. Fritsch and J. Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989) and in T.J. Silhavy, M.L. Berman and L.W. Enquist, Experiments with Gene Fusions, Cold

Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and in Ausubel, F.M. et al., Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley Interscience (1987).

For expression in a suitable host organism, the recombinant nucleic acid construct or gene construct is advantageously inserted into a host-specific vector, which makes optimal expression of the genes in the host possible. Vectors are well known to the skilled worker and can be found, for example, in "Cloning Vectors" (Pouwels P. H. et al., eds, Elsevier, Amsterdam-New York-Oxford, 1985). Vectors also mean not only plasmids but also all other vectors known to the skilled worker, such as, for example, phages, viruses, such as SV40, CMV, baculovirus and adenovirus, transposons, IS elements, phasmids, cosmids, and linear or circular DNA. These vectors may undergo autonomous replication in the host organism or chromosomal replication.

Examples of suitable expression vectors which may be mentioned are:

Conventional fusion expression vectors such as pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) Gene 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT 5 (Pharmacia, Piscataway, NJ), with which respectively glutathione S-transferase (GST), maltose E-binding protein and protein A are fused to the recombinant target protein.

Non-fusion protein expression vectors such as pTrc (Amann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al. Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 60-89).

Yeast expression vector for expression in the yeast *S. cerevisiae*, such as pYepSec1 (Baldari et al., (1987) Embo J. 6:229-234), pMFα (Kurjan and Herskowitz (1982) Cell 30:933-943), pJRY88 (Schultz et al. (1987) Gene 54:113-123) and pYES2 (Invitrogen Corporation, San Diego, CA).

Vectors and methods for constructing vectors suitable for the use in other fungi such as filamentous fungi comprise those which are described in detail in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy et al., eds, pp. 1-28, Cambridge University Press: Cambridge.

Baculovirus vectors which are available for expression of proteins in cultured insect cells (for example Sf9 cells) comprise the pAc series (Smith et al., (1983) Mol. Cell Biol. 3:2156-2165) and pVL series (Lucklow and Summers (1989) Virology 170:31-39).

Further suitable expression systems for prokaryotic and eukaryotic cells are described in chapters 16 and 17 of Sambrook, J., Fritsch, E.F. and Maniatis, T., Molecular cloning: A

Laboratory Manual, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The expression constructs or vectors of the invention can be used to produce genetically modified microorganisms which are transformed, for example, with at least one vector of the invention.

The recombinant constructs of the invention described above are advantageously introduced and expressed in a suitable host system. Cloning and transfection methods familiar to the skilled worker, such as, for example, coprecipitation, protoplast fusion, electroporation, retroviral transfection and the like, are preferably used to bring about expression of said nucleic acids in the particular expression system. Suitable systems are described, for example, in Current Protocols in Molecular Biology, F. Ausubel et al., eds, Wiley Interscience, New York 1997.

Successfully transformed organisms can be selected through marker genes which are likewise present in the vector or in the expression cassette. Examples of such marker genes are genes for antibiotic resistance and for enzymes which catalyze a color-forming reaction which causes staining of the transformed cell. These can then be selected by automatic cell sorting.

Microorganisms which have been successfully transformed with a vector and harbor an appropriate antibiotic resistance gene (for example G418 or hygromycin) can be selected by appropriate antibiotic-containing media or nutrient media. Marker proteins present on the surface of the cell can be used for selection by means of affinity chromatography.

The combination of the host organisms and the vectors appropriate for the organisms, such as plasmids, viruses or phages, such as, for example, plasmids with the RNA polymerase/promoter system, phages 8 or other temperate phages or transposons and/or other advantageous regulatory sequences forms an expression system.

The invention further relates to a process for producing genetically modified organisms, which comprises introducing a nucleic acid construct comprising functionally linked a promoter and nucleic acids encoding a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2, and, where appropriate, a terminator into the genome of the starting organism or extrachromosomally into the starting organism.

The invention further relates to the genetically modified organisms where the genetic modification

A in the case where the wild-type organism already has a ketolase activity, increases the activity of a ketolase compared with the wild type and

B in the case where the wild-type organism has no ketolase activity, causes the activity of a ketolase compared with the wild type,

and the ketolase activity which has been increased as in A or caused as in B is caused by a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2.

As stated above, the increasing or causing of the ketolase activity is brought about by an increasing or causing of the gene expression of a nucleic acid encoding a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2, compared with the wild type.

In a further preferred embodiment, as stated above, the increasing or causing of the gene expression of a nucleic acid encoding a ketolase takes place by introducing nucleic acids encoding a ketolase into the plants and thus preferably by overexpression or transgenic expression of nucleic acids encoding a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2.

The invention further relates to a genetically modified organism comprising at least one transgenic nucleic acid encoding a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2. This is the case when the starting organism has no ketolase or an endogenous ketolase, and a transgenic ketolase is overexpressed.

The invention further relates to a genetically modified organism comprising at least two endogenous nucleic acids encoding a ketolase comprising the amino acid sequence SEQ. ID. NO. 2 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 42% at the amino acid level with the sequence SEQ. ID. NO. 2. This is the case when the starting organism has an endogenous ketolase, and the endogenous ketolase is overexpressed.

Particularly preferred genetically modified organisms have, as mentioned above, additionally an increased hydroxylase activity and/or β-cyclase activity compared with a wild-type organism. Further preferred embodiments are described above in the process of the invention.

Organisms preferably mean according to the invention organisms which are able as wild-type or starting organisms naturally or through genetic complementation and/or reregulation of metabolic pathways to produce carotenoids, in particular β-carotene and/or zeaxanthin and/or neoxanthin and/or violaxanthin and/or luteine.

Further preferred organisms already have as wild-type or starting organisms a hydroxylase activity and are thus able as wild-type or starting organisms to produce zeaxanthin.

Preferred organisms are plants or microorganisms such as, for example, bacteria, yeasts, algae or fungi.

Bacteria which can be used are both bacteria which are able, because of the introduction of genes of carotenoid biosynthesis of a carotenoid-producing organism, to synthesize xanthophylls, such as, for example, bacteria of the genus *Escherichia*, which comprise for example crt genes from *Erwinia*, and bacteria which are intrinsically able to synthesize xanthophylls, such as, for example, bacteria of the genus *Erwinia*, *Agrobacterium*, *Flavobacterium*, *Alcaligenes*, *Paracoccus*, *Nostoc* or cyanobacteria of the genus *Synechocystis*.

Preferred bacteria are *Escherichia coli*, *Erwinia herbicola*, *Erwinia uredovora*, *Agrobacterium aurantiacum*, *Alcaligenes sp. PC-1*, *Flavobacterium* sp. strain R1534, the cyanobacterium *Synechocystis* sp. PCC6803, *Paracoccus marcusii* or *Paracoccus carotinifaciens*.

Preferred yeasts are *Candida*, *Saccharomyces*, *Hansenula*, *Pichia* or *Phaffia*. Particularly preferred yeasts are *Xanthophyllomyces dendrorhous* or *Phaffia rhodozyma*.

Preferred fungi are *Aspergillus*, *Trichoderma*, *Ashbya*, *Neurospora*, *Blakeslea*, *Phycomyces*, *Fusarium* or other fungi described in Indian Chem. Engr. Section B. Vol. 37, No. 1, 2 (1995) on page 15, table 6.

Preferred algae are green algae such as, for example, algae of the genus *Haematococcus*, *Phaedactylum tricornatum*, *Volvox* or *Dunaliella*. Particularly preferred algae are *Haematococcus pluvialis* or *Dunaliella bardawil*.

Further microorganisms which can be used and the production thereof for carrying out the

process of the invention are disclosed for example in DE-A-199 16 140, which is incorporated herein by reference.

Particularly preferred plants are plants selected from the families Ranunculaceae, Berberidaceae, Papaveraceae, Cannabaceae, Rosaceae, Fabaceae, Linaceae, Vitaceae, Brassicaceae, Cucurbitaceae, Primulaceae, Caryophyllaceae, Amaranthaceae, Gentianaceae, Geraniaceae, Caprifoliaceae, Oleaceae, Tropaeolaceae, Solanaceae, Scrophulariaceae, Asteraceae, Liliaceae, Amaryllidaceae, Poaceae, Orchidaceae, Malvaceae, Illiaceae or Lamiaceae.

Very particularly preferred plants are selected from the group of plant genera *Marigold, Tagetes errecta, Tagetes patula, Acacia, Aconitum, Adonis, Arnica, Aquilegia, Aster, Astragalus, Bignonia, Calendula, Caltha, Campanula, Canna, Centaurea, Cheiranthus, Chrysanthemum, Citrus, Crepis, Crocus, Curcurbita, Cytisus, Delonia, Delphinium, Dianthus, Dimorphotheca, Doronicum, Eschscholtzia, Forsythia, Fremontia, Gazania, Gelsemium, Genista, Gentiana, Geranium, Gerbera, Geum, Grevillea, Helenium, Helianthus, Hepatica, Heracleum, Hibiscus, Heliopsis, Hypericum, Hypochoeris, Impatiens, Iris, Jacaranda, Kerria, Laburnum, Lathyrus, Leontodon, Lilium, Linum, Lotus, Lycopersicon, Lysimachia, Maratia, Medicago, Mimulus, Narcissus, Oenothera, Osmanthus, Petunia, Photinia, Physalis, Phyteuma, Potentilla, Pyracantha, Ranunculus, Rhododendron, Rosa, Rudbeckia, Senecio, Silene, Silphium, Sinapsis, Sorbus, Spartium, Tecoma, Torenia, Tragopogon, Trollius, Tropaeolum, Tulipa, Tussilago, Ulex, Viola* or *Zinnia,* particularly preferably selected from the group of plant genera Marigold, Tagetes erecta, Tagetes patula, Lycopersicon, Rosa, Calendula, Physalis, Medicago, Helianthus, Chrysanthemum, Aster, Tulipa, Narcissus, Petunia, Geranium, Tropaeolum or Adonis.

Very particularly preferred genetically modified plants are selected from the plant genera Marigold, Tagetes erecta, Tagetes patula, Adonis, Lycopersicon, Rosa, Calendula, Physalis, Medicago, Helianthus, Chrysanthemum, Aster, Tulipa, Narcissus, Petunia, Geranium or Tropaeolum, with the genetically modified plant comprising at least one transgenic nucleic acid encoding a ketolase.

The present invention further relates to the transgenic plants, their propagation material, and their plant cells, tissues or parts, especially their fruit, seeds, flowers and petals.

The genetically modified plants can, as described above, be used for preparing ketocarotenoids, especially astaxanthin.

Genetically modified organisms of the invention which can be consumed by humans and animals, especially plants or parts of plants, such as, in particular, petals with an increased

content of ketocarotenoids, especially astaxanthin, can also be used directly or after processing known per se as human or animal foods or as animal and human food supplements.

The genetically modified organisms can also be used for producing ketocarotenoid-containing extracts of the organisms and/or for producing animal and human food supplements.

The genetically modified organisms have an increased content of ketocarotenoids compared with the wild type.

An increased content of ketocarotenoids usually means an increased total ketocarotenoid content.

However, an increased content of ketocarotenoid also means in particular an altered content of the preferred ketocarotenoids without the need for the total carotenoid content necessarily to be increased.

In a particularly preferred embodiment, the genetically modified plants of the invention have an increased astaxanthin content compared with the wild type.

An increased content means in this case also a caused content of ketocarotenoids such as astaxanthin.

The invention further relates to the novel ketolases and to the novel nucleic acids which encode the latter.

The invention relates in particular to ketolases comprising the amino acid sequence SEQ. ID. NO. 8 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 70%, preferably at least 75%, particularly preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95% at the amino acid level with the sequence SEQ. ID. NO. 2, with the proviso that the amino acid sequence SEQ. ID NO. 8 is not present. The sequence SEQ ID NO: 4 is, as mentioned above, annotated as putative protein in databases.

The invention further relates to ketolases comprising the amino acid sequence SEQ. ID. NO. 6 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 70% at the amino acid level with the sequence SEQ. ID. NO. 6. The sequence SEQ ID NO: 6 is, as mentioned above, not annotated in databases.

In a further embodiment, the invention relates to ketolases comprising the amino acid sequence SEQ ID NO: 12 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 70%, preferably at least 75%, particularly preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, at the amino acid level with the sequence SEQ ID NO: 12, with the proviso that the amino acid sequence SEQ ID NO: 6 is not present.

The invention further relates to ketolases comprising the amino acid sequence SEQ ID NO: 49 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 50%, preferably at least 60%, particularly preferably at least 70%, more preferably at least 80%, more preferably at least 90%, more preferably at least 95%, at the amino acid level with the sequence SEQ ID NO: 49, with the proviso that the amino acid sequence SEQ ID NO: 47 is not present. The sequence SEQ ID NO: 47 is, as mentioned above, annotated as a putative protein in databases.

The invention further relates to nucleic acids encoding a protein described above, with the proviso that the nucleic acid does not comprise the sequence SEQ ID NO: 5.

It has surprisingly been found that a protein comprising the amino acid sequence SEQ. ID. NO. 4 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 70%, preferably at least 75%, particularly preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, at the amino acid level with the sequence SEQ. ID. NO. 4 and has the property of a ketolase, has a property as ketolase.

The invention therefore also relates to the use of a protein comprising the amino acid sequence SEQ. ID. NO. 4 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 70%, preferably at least 75%, particularly preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, at the amino acid level with the sequence SEQ. ID. NO. 4, and has the property of a ketolase, as ketolase.

It has also surprisingly been found that a protein comprising the amino acid sequence SEQ. ID. NO. 6 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 65%, preferably at least 70 %, preferably at least 75%, particularly preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, at the amino acid level with the sequence SEQ. ID. NO. 6, and has the property of a ketolase, has a property as ketolase.

The invention therefore also relates to the use of a protein comprising the amino acid sequence SEQ. ID. NO. 6 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 65%, preferably at least 70%, preferably at least 75%, particularly preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, at the amino acid level with the sequence SEQ. ID. NO. 6, and has the property of a ketolase, as ketolase.

It has also surprisingly been found that a protein comprising the amino acid sequence SEQ ID NO: 47 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 50%, preferably at least 60%, preferably at least 70%, particularly preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, at the amino acid level with the sequence SEQ ID NO: 47 and which has the property of a ketolase, has a property as ketolase.

The invention therefore also relates to the use of a protein comprising the amino acid sequence SEQ ID NO: 47 or a sequence which is derived from this sequence by substitution, insertion or deletion of amino acids and which has an identity of at least 50%, preferably at least 60%, preferably at least 70%, particularly preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, at the amino acid level with the sequence SEQ ID NO: 47 and which has the property of a ketolase, as ketolase.

Compared with prior art processes, the process of the invention affords a larger quantity of ketocarotenoids, especially astaxanthin having a small quantity of hydroxylated byproducts.

The invention is now explained by the following examples, but is not restricted thereto:

General experimental conditions:
Sequence analysis of recombinant DNA

Recombinant DNA molecules were sequenced using a laser fluorescence DNA sequencer from Licor (sold by MWG Biotech, Ebersbach, Germany), following the method of Sanger (Sanger et al., Proc. Natl. Acad. Sci. USA 74 (1977), 5463-5467).

Example 1:
Amplification of a DNA which encodes the entire primary sequence of NOST ketolase from *Nostoc sp. PCC 7120*

The DNA encoding the *Nostoc sp. PCC 7120* NOST ketolase was amplified from *Nostoc sp. PCC 7120* ("Pasteur Culture Collection of Cyanobacterium" strain) by means of PCR.

To prepare genomic DNA from a *Nostoc sp. PCC 7120* suspension culture which had grown in *BG 11* medium (1.5 g/l $NaNO_3$, 0.04 g/l $K_2PO_4x3H_2O$, 0.075 g/l $MgSO_4xH_2O$, 0.036 g/l $CaCl_2x2H_2O$, 0.006 g/l citric acid, 0.006 g/l ferric ammonium citrate, 0.001 g/l EDTA disodium magnesium, 0.04 g/l Na_2CO_3, 1 ml of trace metal mix A5+Co (2.86 g/l H_3BO_3, 1.81 g/l $MnCl_2x4H_2O$, 0.222 g/l $ZnSO_4x7H_2O$, 0.39 g/l $NaMoO_4x2H_2O$, 0.079 g/l $CuSO_4x5H_2O$, 0.0494 g/l $Co(NO_3)_2x6H_2O$) at 25°C with constant shaking (150 rpm) and under continuous light for 1 week, the cells were harvested by centrifugation, frozen in liquid nitrogen and ground to a powder in a mortar.

Protocol for isolating DNA from Nostoc PCC7120:

The bacteria cells were pelleted from a 10 ml liquid culture by centrifugation at 8000 rpm for 10 minutes. The bacterial cells were then crushed and ground in liquid nitrogen, using a mortar. The cell material was resuspended in 1 ml of 10 mM Tris HCl (pH 7.5) and transferred to an Eppendorf reaction vessel (volume: 2 ml). After addition of 100 µl of proteinase K (concentration: 20 mg/ml), the cell suspension was incubated at 37°C for 3 hours. The suspension was then extracted with 500 µl of phenol. After centrifugation at 13 000 rpm for 5 minutes, the upper, aqueous phase was transferred to a new 2 ml Eppendorf reaction vessel. Extraction with phenol was repeated 3 times. The DNA was precipitated by adding 1/10 volume of 3 M sodium acetate (pH 5.2) and 0.6 volume of isopropanol and then washed with 70% ethanol. The DNA pellet was dried at room temperature, taken up in 25 µl of water and dissolved with heating to 65°C.

The nucleic acid encoding a *Nostoc PCC 7120* ketolase was amplified from *Nostoc sp. PCC 7120* by means of polymerase chain reaction (PCR) using a sense-specific primer (NOSTF, SEQ ID NO. 19) and an antisense-specific primer (NOSTG SEQ ID NO. 20).

The PCR conditions were as follows:

The PCR for amplifying the DNA encoding a ketolase protein consisting of the entire primary sequence was carried out in a 50 µl reaction mixture which contained:

- 1 µl of a *Nostoc sp. PCC 7120* DNA (prepared as described above)
- 0.25 mM dNTPs
- 0.2 mM NOSTF (SEQ ID NO. 19)
- 0.2 mM NOSTG (SEQ ID NO. 20)
- 5 µl of 10X PCR buffer (TAKARA)
- 0.25 µl of R Taq polymerase (TAKARA)
- 25.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
35X	94°C for 1 minute
	55°C for 1 minute
	72°C for 3 minutes
1X	72°C for 10 minutes

PCR amplification with SEQ ID NO. 19 and SEQ ID NO. 20 resulted in an 805 bp fragment encoding a protein consisting of the entire primary sequence (SEQ ID NO. 21). Using standard methods, the amplicon was cloned into the PCR cloning vector pGEM-T (Promega), producing the clone pNOSTF-G.

Sequencing of the pNOSTF-G clone with the M13F and M13R primers confirmed a sequence which is identical to the DNA sequence from 88,886-89,662 of database entry AP003592. This nucleotide sequence was reproduced in an independent amplification experiment and thus represents the nucleotide sequence in the *Nostoc sp. PCC 7120* used.

Therefore, said clone, pNOSTF-G, was used for cloning into the expression vector pJIT117 (Guerineau et al. 1988, Nucl. Acids Res. 16: 11380). Cloning was carried out by isolating the 799 bp SphI fragment from pNOSTF-G and ligating it into the SphI-cut pJIT117 vector. The clone which contains the *Nostoc sp. PCC 7120* ketolase in the correct orientation as an N-terminal translational fusion with the rbcS transit peptide is referred to as pJNOST.

Example 2:
Construction of the plasmid pMCL-CrtYIBZ/idi/gps for the synthesis of zeaxanthin in *E. coli*

pMCL-CrtYIBZ/idi/gps was constructed in three steps via the intermediates pMCL-CrtYIBZ and pMCL-CrtYIBZ/idi. The vector used was the plasmid pMCL200 which is compatible with high copy-number vectors (Nakano, Y., Yoshida, Y., Yamashita, Y. and Koga, T.; Construction of a series of pACYC-derived plasmid vectors; Gene 162 (1995), 157-158).

Example 2.1.: Construction of pMCL-CrtYIBZ

The biosynthesis genes *crtY, crtB, crtI* and *crtZ* are from the bacterium *Erwinia uredovora* and were amplified by means of PCR. *Erwinia uredovora* (DSM 30080) genomic DNA was prepared by the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Brunswick, Germany) as part of a service. The PCR was carried out according to the manufacturer's information (Roche, Long Template PCR: Procedure for amplification of 5-20 kb targets with the

Expand Long Template PCR system). The PCR conditions for amplifying the biosynthesis cluster of *Erwinia uredovora* were as follows

Master Mix 1:

- 1.75 µl of dNTPs (final concentration 350 µM)
- 0.3 µM primer Crt1 (SEQ ID NO. 22)
- 0.3 µM primer Crt2 (SEQ ID NO. 23)
- 250-500 ng of DSM 30080 genomic DNA

Distilled water to a total volume of 50 µl

Master Mix 2:

- 5 µl of 10x PCR buffer 1 (final concentration 1x, with 1.75 mM Mg2+)
- 10x PCR buffer 2 (final concentration 1x, with 2.25 mM Mg2+)
- 10x PCR buffer 3 (final concentration 1x, with 2.25 mM Mg2+)
- 0.75 µl of Expand Long Template Enzyme Mix (final concentration 2.6 units)

Distilled water to a total volume of 50 µl

The two mixtures "Master Mix 1" and "Master Mix 2" were combined by pipetting. The PCR was carried out in a total volume of 50 µl under the following cycle conditions:

1X	94°C for 2 minutes
30X	94°C for 30 seconds
	58°C for 1 minute
	68°C for 4 minutes
1X	72°C for 10 minutes

PCR amplification with SEQ ID NO. 22 and SEQ ID NO. 23 resulted in a fragment (SEQ ID NO. 24) encoding the genes *CrtY* (protein: SEQ ID NO. 25), *CrtI* (protein: SEQ ID NO. 26), *crtB* (protein: SEQ ID NO. 27) and *CrtZ (iDNA)*. Using standard methods, the amplicon was cloned into the PCR cloning vector pCR2.1 (Invitrogen), producing the clone pCR2.1-CrtYIBZ.

The pCR2.1-CrtYIBZ plasmid was cut with SalI and HindIII, the resulting SalI/HindIII fragment was isolated and transferred by way of ligation into the SalI/HindIII-cut vector pMCL200. The pCR2.1-CrtYIBZ SalI/HindIII fragment cloned into pMCL 200 is 4624 bp in length, encodes the genes *CrtY, CrtI, crtB* and *CrtZ* and corresponds to the sequence from position 2295 to position 6918 in D90087 (SEQ ID NO. 24). The resulting clone is referred to as pMCL-CrtYIBZ.

Example 2.2.: Construction of pMCL-CrtYIBZ/idi

The gene *idi* (isopentenyl-diphosphate isomerase; IPP isomerase) was amplified from *E. coli* by means of PCR. The nucleic acid which encodes the entire *idi* gene including the *idi* promoter and ribosomal binding site was amplified from *E. coli* by means of polymerase chain reaction (PCR) using a sense-specific primer (5'-idi SEQ ID NO. 28) and an antisense-specific primer (3'-idi SEQ ID NO. 29).

The PCR conditions were as follows:

The PCR for amplifying the DNA was carried out in a 50 µl reaction mixture which contained:

- 1 µl of an *E. coli* TOP10 suspension
- 0.25 mM dNTPs
- 0.2 mM 5'-idi (SEQ ID NO. 28)
- 0.2 mM 3'-idi (SEQ ID NO. 29)
- 5 µl of 10X PCR buffer (TAKARA)
- 0.25 µl of R Taq polymerase (TAKARA)
- 28.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
20X	94°C for 1 minute
	62°C for 1 minute
	72°C for 1 minute
1X	72°C for 10 minutes

PCR amplification with SEQ ID NO. 28 and SEQ ID NO. 29 resulted in a 679 bp fragment encoding a protein consisting of the entire primary sequence (SEQ ID NO. 30). Using standard methods, the amplicon was cloned into the PCR cloning vector pCR2.1 (Invitrogen), producing the clone pCR2.1-idi.

Sequencing of the pCR2.1-idi clone confirmed a sequence which does not differ from the published sequence AE000372 in positions 8774 to 9440. This region comprises the promoter region, the potential ribosomal binding site and the entire IPP isomerase open reading frame. The fragment cloned into pCR2.1-idi has a total length of 679 bp, due to insertion of an XhoI cleavage site at the 5' end and an SalI cleavage site at the 3' end of the *idi* gene.

This clone was therefore used for cloning the *idi* gene into the pMCL-CrtYIBZ vector. Cloning

was carried out by isolating the XhoI/SalI fragment from pCR2.1-idi and ligating it into the XhoI/SalI-cut pMCL-CrtYIBZ vector. The resulting clone is referred to as pMCL-CrtYIBZ/idi.

Example 2.3.: Construction of pMCL-CrtYIBZ/idi/gps

The gene *gps* (geranylgeranyl-pyrophosphate synthase; GGPP synthase) was amplified from *Archaeoglobus fulgidus* by means of PCR. The nucleic acid encoding *Archaeoglobus fulgidus gps* was amplified by means of polymerase chain reaction (PCR) using a sense-specific primer (5'-gps SEQ ID NO. 32) and an antisense-specific primer (3'-gps SEQ ID NO. 33).

The *Archaeoglobus fulgidus* DNA was prepared by the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Brunswick, Germany) as part of a service. The PCR conditions were as follows:

The PCR for amplifying the DNA encoding a GGPP synthase protein consisting of the entire primary sequence was carried out in a 50 µl reaction mixture which contained:

- 1 µl of an *Archaeoglobus fulgidus* DNA
- 0.25 mM dNTPs
- 0.2 mM 5'-gps (SEQ ID NO. 32)
- 0.2 mM 3'-gps (SEQ ID NO. 33)
- 5 µl of 10X PCR buffer (TAKARA)
- 0.25 µl of R Taq polymerase (TAKARA)
- 28.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
20X	94°C for 1 minute
	56°C for 1 minute
	72°C for 1 minute
1X	72°C for 10 minutes

The DNA fragment amplified by means of PCR and the primers SEQ ID NO. 32 and SEQ ID NO. 33 was eluted from the agarose gel by methods known per se and cut with the restriction enzymes NcoI and HindIII. This resulted in a 962 bp fragment which encodes a protein consisting of the entire primary sequence (SEQ ID NO. 34). Using standard methods, the NcoI/HindIII-cut amplicon was cloned into the pCB97-30 vector, producing the clone pCB-gps.

Sequencing of the pCB-gps clone confirmed a sequence for *A. fulgidus* GGPP synthase, which

differs from the published sequence AF120272 in one nucleotide. Introducing an NcoI cleavage site in the *gps* gene altered the second codon of GGPP synthase. In the published sequence, AF120272, CTG (positions 4-6) codes for leucine. Amplification with the two primers SEQ ID NO. 32 and SEQ ID NO. 33 altered this second codon to GTG which codes for valine.

The clone pCB-gps was therefore used for cloning the *gps* gene into the pMCL-CrtYIBZ/idi vector. Cloning was carried out by isolating the KpnI/XhoI fragment from pCB-gps and ligating it into the pMCL-CrtYIBZ/idi vector cut with KpnI and XhoI. The cloned KpnI/XhoI fragment (SEQ ID NO. 34) carries the Prrn16 promoter together with a minimum 5' UTR sequence of rbcL, the first 6 rbcL codons which extend the GGPP synthase N-terminally, and, 3' from the *gps* gene, the psbA sequence. Thus, the N terminus of GGPP synthase has, instead of the natural amino acid sequence with Met-Leu-Lys-Glu (amino acids 1 to 4 of AF120272), the altered amino acid sequence Met-Thr-Pro-Gln-Thr-Ala-Met-Val-Lys-Glu. This leads to recombinant GGPP synthase, starting with Lys at position 3 (in AF120272), being identical and having no other changes in the amino acid sequence. The rbcL and psbA sequences were used according to a reference by Eibl et al. (Plant J. 19. (1999), 1-13). The resulting clone is referred to as pMCL-CrtYIBZ/idi/gps.

Example 3:

Biotransformation of zeaxanthin in recombinant E. coli strains

Zeaxanthin biotransformation was carried out by preparing recombinant *E. coli* strains which are capable of zeaxanthin production due to heterologous complementation. *E. coli* TOP10 strains were used as host cells for complementation experiments with the plasmids pNOSTF-G and pMCL-CrtYIBZ/idi/gps.

In order to prepare *E. coli* strains which enable zeaxanthin to be synthesized at high concentrations, plasmid pMCL-CrtYIBZ/idi/gps was constructed. Said plasmid carries the biosynthesis genes *crtY, crtB, crtI* and *crtY* of *Erwinia uredovora*, the *Archaeoglobus fulgidus* gene *gps* (for geranylgeranyl-pyrophosphate synthatase) and the *E. coli* gene *idi* (isopentenyl-diphosphate isomerase). This construct was used to eliminate steps which limit high accumulation of carotenoids and of their biosynthetic precursors. This has been described previously by Wang et al. in a similar manner, using several plasmids (Wang, C.-W., Oh, M.-K. and Liao, J.C.; Engineered isoprenoid pathway enhances astaxanthin production in *Escherichia coli*, Biotechnology and Bioengineering 62 (1999), 235-241).

E. coli TOP10 cultures were transformed in a manner known per se with the two plasmids pNOSTF-G and pMCL-CrtYIBZ/idi/gps and cultured in LB medium at 30°C and 37°C, respectively, overnight. Ampicillin (50 µg/ml), chloramphenicol (50 µg/ml) and isopropyl-β-thio-

galactoside (1 mmol) were likewise added in a manner known per se overnight.

The carotenoids were isolated from the recombinant strains by extracting the cells with acetone, evaporating the organic solvent to dryness and fractionating said carotenoids by means of HPLC via a C30 column. The following process conditions were set.

Separating column: Prontosil C30 column, 250 × 4.6 mm (Bischoff, Leonberg, Germany)
Flow rate: 1.0 ml/min
Eluents: Eluent A - 100% methanol
Eluent B - 80% methanol, 0.2% ammonium acetate
Eluent C - 100% t-butyl methyl ether

Gradient profile:

Time	Flow rate	% eluent A	% eluent B	% eluent C
1.00	1.0	95.0	5.0	0
1.05	1.0	80.0	5.0	15.0
14.00	1.0	42.0	5.0	53.0
14.05	1.0	95.0	5.0	0
17.00	1.0	95.0	5.0	0
18.00	1.0	95.0	5.0	0

Detection: 300-500 nm

The spectra were determined directly from the elution peaks, using a photodiode array detector. The isolated substances were identified by way of their absorption spectra and their retention times, in comparison with standard samples.

Figure 1 depicts the chromatographic analysis of a sample obtained from an *E. coli* strain transformed with pNOSTF-G and pMCL-CrtYIBZ/idi/gps. This strain is shown to be able to synthesize various ketocarotenoids, owing to heterologous complementation. Astaxanthin (peak 1), adonirubin (peak 2) and canthaxanthin (peak 3) are eluted with increasing retention time.

Example 3.1
Comparative example

An *E. coli* strain expressing a ketolase from *Haematococcus pluvialis Flotow em. Wille* was prepared as a comparative example, similarly to the preceding examples. For this purpose, the cDNA encoding the entire primary sequence of *Haematococcus pluvialis Flotow em. Wille* ketolase was amplified and cloned according to example 1 into the same expression vector.

The cDNA encoding *Haematococcus pluvialis* ketolase was amplified from Haematococcus pluvialis (strain 192.80 of the "Sammlung von Algenkulturen der Universität Göttingen") suspension culture by means of PCR. To prepare total RNA from a Haematococcus pluvialis (strain 192.80) suspension culture which had grown in Haematococcus medium (1.2 g/l sodium acetate, 2 g/l yeast extract, 0.2 g/l MgCl2x6H2O, 0.02 CaCl2x2H2O; pH 6.8; after autoclaving, addition of 400 mg/l L-asparagine, 10 mg/l FeSO4xH2O) at room temperature with indirect daylight for 2 weeks, the cells were harvested, frozen in liquid nitrogen and ground to a powder in a mortar. Subsequently, 100 mg of the frozen algal cell powder were transferred to a reaction vessel and taken up in 0.8 ml of Trizol buffer (Life Technologies). The suspension was extracted with 0.2 ml of chloroform. After centrifugation at 12 000 g for 15 minutes, the aqueous supernatant was removed and transferred to a new reaction vessel and extracted with one volume of ethanol. The RNA was precipitated with one volume of isopropanol, washed with 75% ethanol and the pellet was dissolved in DEPC water (water incubated overnight with 1/1000 volume of diethyl pyrocarbonate at room temperature, then autoclaved). The RNA concentration was determined photometrically.

For cDNA synthesis, 2.5 µg of total RNA were denatured at 60°C for 10 min, cooled on ice for 2 min and transcribed into cDNA by means of a cDNA kit (Ready-to-go-you-prime-beads, Pharmacia Biotech), according to the manufacturer's information using an antisense-specific primer, PR1 (gcaagctcga cagctacaaa cc).

The nucleic acid encoding a ketolase from Haematococcus pluvialis (strain 192.80) was amplified by means of polymerase chain reaction (PCR) from Haematococcus pluvialis, using a sense-specific primer, PR2 (gaagcatgca gctagcagcg acag), and an antisense-specific primer, PR1.

The PCR conditions were as follows:

The PCR for amplifying the cDNA encoding a ketolase protein consisting of the entire primary sequence was carried out in a 50 ml reaction mixture containing:

- 4 ml of a Haematococcus pluvialis cDNA (prepared as described above)
- 0.25 mM dNTPs
- 0.2 mM PR1
- 0.2 mM PR2
- 5 ml of 10X PCR buffer (TAKARA)
- 0.25 ml of R Taq polymerase (TAKARA)
- 28.8 ml of distilled water

The PCR was carried out under the following cycle conditions:

	1X	94°C for 2 minutes
5	35X	94°C for 1 minute
		53°C for 2 minutes
		72°C for 3 minutes
	1X	72°C for 10 minutes

10 PCR amplification with PR1 and PR2 resulted in a 1155 bp fragment which encodes a protein consisting of the entire primary sequence:

```
gaagcatgca gctagcagcg acagtaatgt tggagcagct taccggaagc gctgaggcac      60
tcaaggagaa ggagaaggag gttgcaggca gctctgacgt gttgcgtaca tgggcgaccc     120
agtactcgct tccgtcagag gagtcagacg cggcccgccc gggactgaag aatgcctaca     180
agccaccacc ttccgacaca aagggcatca caatggcgct agctgtcatc ggctcctggg     240
ccgcagtgtt cctccacgcc atttttcaaa tcaagcttcc gacctccttg gaccagctgc     300
actggctgcc cgtgtcagat gccacagctc agctggttag cggcagcagc agcctgctgc     360
acatcgtcgt agtattcttt gtcctggagt tcctgtacac aggccttttt atcaccacgc     420
atgatgctat gcatggcacc atcgccatga gaaacaggca gcttaatgac ttcttgggca     480
gagtatgcat ctccttgtac gcctggtttg attacaacat gctgcaccgc aagcattggg     540
agcaccacaa ccacactggc gaggtgggca aggaccctga cttccacagg ggaaaccctg     600
gcattgtgcc ctggtttgcc agcttcatgt ccagctacat gtcgatgtgg cagtttgcgc     660
gcctcgcatg gtggacggtg gtcatgcagc tgctgggtgc gccaatggcg aacctgctgg     720

tgttcatggc ggccgcgccc atcctgtccg ccttccgctt gttctacttt ggcacgtaca     780
tgccccacaa gcctgagcct ggcgccgcgt caggctcttc accagccgtc atgaactggt     840
ggaagtcgcg cactagccag gcgtccgacc tggtcagctt tctgacctgc taccacttcg     900
acctgcactg ggagcaccac cgctggccct ttgccccctg gtgggagctg cccaactgcc     960
gccgcctgtc tggccgaggt ctggttcctg cctagctgga cacactgcag tgggccctgc    1020
tgccagctgg gcatgcaggt tgtggcagga ctgggtgagg tgaaaagctg caggcgctgc    1080
tgccggacac gctgcatggg ctaccctgtg tagctgccgc cactagggga gggggtttgt    1140
agctgtcgag cttgc
```

15

Using standard methods, the amplicon was cloned into the PCR cloning vector pGEM-Teasy (Promega), producing the clone pGKETO2.

Sequencing of the pGKETO2 clone, using the T7 and SP6 primers, confirmed a sequence which
20 differs from the published sequence, X86782, only in the three codons 73, 114 and 119 by one base each. These nucleotide substitutions were reproduced in an independent amplification experiment and thus represent the nucleotide sequence in the Haematococcus pluvialis strain used, 192.80.

25 This clone was used for cloning into the expression vector described in example 1. Cloning was carried out in a manner similar to that described in example 1. Transformation of the E. coli

strains, culturing thereof and analysis of the carotenoid profile were carried out as described in example 3.

Figure 2 depicts the chromatographic analysis of a sample obtained from an *E. coli* strain transformed with said expression vector and pMCL-CrtYIBZ/idi/gps. With use of a *Haematococcus pluvialis* ketolase, as described, for example, in EP 725137, astaxanthin (peak 1), adonixanthin (peak 2) and unreacted zeaxanthin (peak 3) elute with increasing retention time. This carotenoid profile has already been described in EP 0725137.

Table 1 depicts a comparison of the bacterially produced carotenoid quantities:

Table 1: Comparison of bacterial ketocarotenoid synthesis using two different ketolases, the *Nostoc* sp. PCC7120 NOST ketolase according to the invention (example 3) and *Haematococcus pluvialis* ketolase as a comparative example (example 3.1). Carotenoid quantities are indicated in ng/ml culture liquid.

Ketolase from	Astaxanthin	Adonirubin	Adonixanthin	Canthaxanthin	Zeaxanthin
Haematococcus pluvialis Flotow em. Wille (comparative example)	13		102		738
Nostoc sp. Strain PCC7120	491	186		120	

Expression of *Nostoc sp.* strain PCC7120 ketolase according to the invention results in a carotenoid pattern which differs markedly from the carotenoid pattern after expression of a *Haematococcus pluvialis* ketolase. While the ketolase of the prior art provides the desired ketocarotenoid astaxanthin only in very limited amounts, astaxanthin is the main product when using the ketolase according to the invention. A distinctly lower amount of hydroxylated byproducts appears in the process of the invention.

Example 4:
Preparation of expression vectors for constitutive expression of *Nostoc sp. PCC7120* NOST ketolase in *Lycopersicon esculentum* and *Tagetes erecta*.

The *Nostoc sp. PCC7120* NOST ketolase is expressed in *L. esculentum* and in *Tagetes erecta* under the control of the constitutive promoter FNR (ferredoxin NADPH oxidoreductase, database entry AB011474, positions 70127 to 69493; WO03/006660) from *Arabidopsis thaliana*. The FNR gene starts at base pair 69492 and is annotated with "ferredoxin-NADP+ reductase".

Expression was carried out using the pea transit peptide rbcS (Anderson et al. 1986, Biochem J. 240: 709-715).

The DNA fragment comprising the *Arabidopsis thaliana* FNR promoter region was prepared by means of PCR using genomic DNA (isolated from *Arabidopsis thaliana* by standard methods) and the primers FNR-A (SEQ ID NO. 38) and FNR-B (SEQ ID NO. 39).

The PCR conditions were as follows:

The PCR for amplifying the DNA comprising the FNR promoter fragment FNR#1) was carried out in a 50 µl reaction mixture containing:

- 100 ng of *A. thaliana* genomic DNA
- 0.25 mM dNTPs
- 0.2 mM FNR-A (SEQ ID NO. 38)
- 0.2 mM FNR-B (SEQ ID NO. 39)
- 5 µl of 10X PCR buffer (Stratagene)
- 0.25 µl of Pfu polymerase (Stratagene)
- 28.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
35X	94°C for 1 minute
	50°C for 1 minute
	72°C for 1 minute
1X	72°C for 10 minutes

The 647 bp amplicon was cloned into the PCR cloning vector PCR 2.1 (Invitrogen) by using standard methods, producing the plasmid pFNR#1.

Sequencing of the pFNR#1 clone confirmed a sequence which corresponds to a sequence section on chromosome 5 of Arabidopsis thaliana (database entry AB011474; WO03/006660), from position 70127 to position 69493. The FNR gene starts at base pair 69492 and is annotated with "ferredoxin-NADP+ reductase".

pFNR was therefore used for cloning into the expression vector pJIT117 (Guerineau et al. 1988, Nucl. Acids Res. 16: 11380).

Cloning was carried out by isolating the 637 bp SacI-HindIII fragment from pFNR#1 (partial SacI hydrolysis) and ligating it into the SacI-HindIII-cut pJIT117 vector. The clone which contains the FNR#1 promoter instead of the original d35S promoter is referred to as pJITFNR.

An expression cassette, pJFNRNOST, was prepared by cloning the 799 bp SpHI fragment, NOSTF-G (described in example 1), into the SpHI-cut pJITFNR vector. The clone which contains the NOSTF-G fragment in the correct orientation as N-terminal fusion with the rbcS transit peptide is referred to as pJFNRNOST.

An expression cassette for Agrobacterium-mediated transformation of *Nostoc* ketolase into *L. esculentum* was prepared using the binary vector pSUN3 (WO02/00900).

The expression vector pS3FNR:NOST (MSP101) was prepared by ligating the 2.425 bp SacI-XhoI fragment (partial SacI hydrolysis) from pJFNRNOST with the SacI-XhoI-cut pSUN3 vector (figure 3, construct map). In figure 3, the fragment *FNR promoter* comprises the FNR promoter (635 bp), the fragment *rbcS TP fragment* comprises the pea rbcS transit peptide (194 bp), the fragment *Nost ketolase CDS* (777 bp) comprises the entire primary sequence coding for *Nostoc* ketolase, and the fragment *35S Term* (746 bp) comprises the CaMV polyadenylation signal.

An expression cassette for *Agrobacterium*-mediated transformation of the expression vector containing *Nostoc* ketolase into Tagetes erecta was prepared using the binary vector pSUN5 (WO02/00900).

The Tagetes expression vector pS5FNR:NOST (MSP102) was prepared by ligating the 2.425 bp SacI-XhoI fragment (partial SacI hydrolysis) from pJFNRNOST with the SacI-XhoI-cut pSUN5 vector (figure 4, construct map). In figure 4, the fragment *FNR promoter* comprises the FNR promoter (635 bp), the fragment *rbcS Transit Peptide* comprises the pea rbcS transit peptide (194 bp), the fragment *Nost ketolase* (777 bp) comprises the entire primary sequence coding for *Nostoc* ketolase, and the fragment *35S Terminator* (746 bp) comprises the CaMV polyadenylation signal.

Example 5:

Preparation of expression vectors for flower-specific expression of *Nostoc sp. PCC 7120 NOST* ketolase in *Lycopersicon esculentum* and *Tagetes erecta*.

Expression of *Nostoc* ketolase in L. esculentum and Tagetes erecta was carried out using the pea transit peptide rbcS (Anderson et al. 1986, Biochem J. 240: 709-715). Expression was carried out under the control of a modified version, AP3P, of the flower-specific Arabidopsis thaliana promoter AP3 (AL132971: nucleotide region 9298-10200; Hill et al. (1998) Development

125: 1711-1721).

The DNA fragment which comprises the AP3 promoter region -902 to +15 from Arabidopsis thaliana was prepared by means of PCR using genomic DNA (isolated from Arabidopsis thaliana by standard methods) and the primers AP3-1 (SEQ ID NO. 41) and AP3-2 (SEQ ID NO. 42).

The PCR conditions were as follows:

The PCR for amplifying the DNA comprising the AP3 promoter fragment (-902 to +15) was carried out in a 50 µl reaction mixture containing:

- 100 ng of A. thaliana genomic DNA
- 0.25 mM dNTPs
- 0.2 mM AP3-1 (SEQ ID NO. 41)
- 0.2 mM AP3-2 (SEQ ID NO. 42)
- 5 µl of 10X PCR buffer (Stratagene)
- 0.25 µl of Pfu polymerase (Stratagene)
- 28.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
35X	94°C for 1 minute
	50°C for 1 minute
	72°C for 1 minute
1X	72°C for 10 minutes

The 929 bp amplicon was cloned into the PCR cloning vector PCR 2.1 (Invitrogen) by using standard methods, producing the plasmid pAP3.

Sequencing of the pAP3 clone confirmed a sequence which differs from the published AP3 sequence (AL132971, nucleotide region 9298-10200) only by an insertion (a G in position 9765 of the sequence AL132971) and a base substitution (G for A in position 9726 of the sequence AL132971). These nucleotide differences were reproduced in an independent amplification experiment and thus represent the actual nucleotide sequence in the Arabidopsis thaliana plants used.

The modified version, AP3P, was prepared by means of recombinant PCR using the pAP3 plasmid. The region 10200-9771 was amplified using the primers AP3-1 (SEQ ID NO. 41) and

primers AP3-4 (SEQ ID NO. 44) (amplicon A1/4), and the region 9526-9285 was amplified using AP3-3 (SEQ ID NO. 43) and AP3-2 (SEQ ID NO. 42) (amplicon A2/3).

The PCR conditions were as follows:

The PCRs for amplifying the DNA fragments comprising the regions region 10200-9771 and region 9526-9285 of the AP3 promoter were carried out in 50 µl reaction mixtures containing:

- 100 ng of AP3 amplicon (described above)
- 0.25 mM dNTPs
- 0.2 mM sense primer (AP3-1 SEQ ID NO. 41 or AP3-3 SEQ ID NO. 43)
- 0.2 mM antisense primer (AP3-4 SEQ ID NO. 44 or AP3-2 SEQ ID NO. 42)
- 5 µl of 10X PCR buffer (Stratagene)
- 0.25 µl of Pfu Taq polymerase (Stratagene)
- 28.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
35X	94°C for 1 minute
	50°C for 1 minute
	72°C for 1 minute
1X	72°C for 10 minutes

The recombinant PCR comprises annealing of the amplicons A1/4 and A2/3 which overlap over a sequence of 25 nucleotides, completion to give a double strand and subsequent amplification. This results in a modified version of the AP3 promoter, AP3P, in which positions 9670-9526 have been deleted.

The two amplicons A1/4 and A2/3 were denatured (5 min at 95°C) and annealed (slow cooling to 40°C at room temperature) in a 17.6 µl reaction mixture which contained:

- 0.5 µg of A1/4 amplicon
- 0.25 µg of A2/3 amplicon

The 3' ends were filled in (30 min at 30°C) in a 20 µl reaction mixture which contained:

- 17.6 µl of A1/4 and A2/3 annealing reaction (prepared as described above)
- 50 µm dNTPs

- 2 µl of 1X Klenow buffer
- 2U of Klenow enzyme

The nucleic acid coding for the modified promoter version, AP3P, was amplified by means of PCR using a sense-specific primer (AP3-1 SEQ ID NO. 41) and an antisense-specific primer (AP3-2 SEQ ID NO. 42).

The PCR conditions were as follows:

The PCR for amplifying the AP3P fragment was carried out in a 50 µl reaction mixture containing:

- 1 µl of annealing reaction (prepared as described above)
- 0.25 mM dNTPs
- 0.2 mM AP3-1 (SEQ ID NO. 41)
- 0.2 mM AP3-2 (SEQ ID NO. 42)
- 5 µl of 10X PCR buffer (Stratagene)
- 0.25 µl of Pfu Taq polymerase (Stratagene)
- 28.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
35X	94°C for 1 minute
	50°C for 1 minute
	72°C for 1 minute
1X	72°C for 10 minutes

PCR amplification with SEQ ID NO. 41 (AP3-1) and SEQ ID NO. 42 (AP3-2) resulted in a 777 bp fragment encoding the modified promoter version, AP3P. The amplicon was cloned into the cloning vector pCR2.1 (Invitrogen), producing the plasmid pAP3P. Sequencing reactions using the primers T7 and M13 confirmed a sequence identical to the sequence AL132971, region 10200-9298, with the internal region 9285-9526 having been deleted. This clone was therefore used for cloning into the expression vector pJIT117 (Guerineau et al. 1988, Nucl. Acids Res. 16: 11380).

Cloning was carried out by isolating the 767 bp SacI-HindIII fragment from pAP3P and ligating it into the SacI-HindIII-cut pJIT117 vector. The clone which contains the AP3P promoter instead of the original d35S promoter is referred to as pJITAP3P. An expression cassette, pJAP3NOST,

was prepared by cloning the 799 bp SpHI fragment, NOSTF-G (described in example 1), into the SpHI-cut pJITAP3P vector. The clone which contains the NOSTF-G fragment in the correct orientation as an N-terminal fusion with the rbcS transit peptide is referred to as pJAP3PNOST.

An expression vector for Agrobacterium-mediated transformation of the AP3P-controlled Nostoc ketolase into *L. esculentum* was prepared using the binary vector pSUN3 (WO02/00900).

The expression vector pS3AP3:NOST (MSP103) was prepared by ligating the 2.555 bp SacI-XhoI fragment from pJAP3NOST with the SacI-XhoI-cut pSUN3 vector (figure 5, construct map). In figure 5, the fragment AP3P PROMOTER comprises the modified AP3P promoter (765 bp), the fragment rbcS TP FRAGMENT comprises the pea rbcS transit peptide (194 bp), the fragment NOST KETOLASE CDS (777 bp) comprises the entire primary sequence coding for Nostoc ketolase, and the fragment 35S TERM (746 bp) comprises the CaMV polyadenylation signal.

An expression vector for Agrobacterium-mediated transformation of the AP3P-controlled Nostoc ketolase into Tagetes erecta was prepared using the binary vector pSUN5 (WO02/00900).

The expression vector pS5AP3:NOST (MSP104) was prepared by ligating the 2.555 bp SacI-XhoI fragment from pS5AP3PNOST with the SacI-XhoI-cut pSUN5 vector (figure 6, construct map). In figure 6, the fragment AP3P PROMOTER comprises the modified AP3P promoter (765 bp), the fragment rbcS TP FRAGMENT comprises the pea rbcS transit peptide (207 bp), the fragment NOST KETOLASE CDS (777 bp) comprises the entire primary sequence coding for Nostoc ketolase, and the fragment 35S TERM (746 bp) comprises the CaMV polyadenylation signal.

Example 6

Amplification of a DNA encoding the entire primary sequence of NP196 ketolase from Nostoc *punctiforme ATCC 29133*.

The DNA encoding the *Nostoc punctiforme ATCC 29133* NP196 ketolase was amplified from *Nostoc punctiforme ATCC 29133* ("American Type Culture Collection" strain) by means of PCR.

To prepare genomic DNA from a *Nostoc punctiforme ATCC 29133* suspension culture which had grown in *BG 11* medium (1.5 g/l $NaNO_3$, 0.04 g/l $K_2PO_4x3H_2O$, 0.075 g/l $MgSO_4xH_2O$, 0.036 g/l $CaCl_2x2H_2O$, 0.006 g/l citric acid, 0.006 g/l ferric ammonium citrate, 0.001 g/l EDTA disodium magnesium, 0.04 g/l Na_2CO_3, 1 ml of trace metal mix A5+Co (2.86 g/l H_3BO_3, 1.81 g/l $MnCl_2x4H_2O$, 0.222 g/l $ZnSO_4x7H_2O$, 0.39 g/l $NaMoO_4x2H_2O$, 0.079 g/l $CuSO_4x5H_2O$, 0.0494 g/l $Co(NO_3)2x6H_2O$) at 25°C with constant shaking (150 rpm) and under continuous light for 1

week, the cells were harvested by centrifugation, frozen in liquid nitrogen and ground to a powder in a mortar.

Protocol for isolating DNA from *Nostoc punctiforme ATCC 29133*:

The bacteria cells were pelleted from a 10 ml liquid culture by centrifugation at 8000 rpm for 10 minutes. The bacterial cells were then crushed and ground in liquid nitrogen, using a mortar. The cell material was resuspended in 1 ml of 10 mM Tris HCl (pH 7.5) and transferred to an Eppendorf reaction vessel (volume: 2 ml). After addition of 100 µl of proteinase K (concentration: 20 mg/ml), the cell suspension was incubated at 37°C for 3 hours. The suspension was then extracted with 500 µl of phenol. After centrifugation at 13 000 rpm for 5 minutes, the upper, aqueous phase was transferred to a new 2 ml Eppendorf reaction vessel. Extraction with phenol was repeated 3 times. The DNA was precipitated by adding 1/10 volume of 3 M sodium acetate (pH 5.2) and 0.6 volume of isopropanol and then washed with 70% ethanol. The DNA pellet was dried at room temperature, taken up in 25 µl of water and dissolved with heating to 65°C.

The nucleic acid encoding a *Nostoc punctiforme ATCC 29133* ketolase was amplified from *Nostoc punctiforme ATCC 29133* by means of polymerase chain reaction (PCR) using a sense-specific primer (NP196-1, SEQ ID NO. 54) and an antisense-specific primer (NP196-2 SEQ ID NO. 55).

The PCR conditions were as follows:

The PCR for amplifying the DNA encoding a ketolase protein consisting of the entire primary sequence was carried out in a 50 µl reaction mixture which contained:

- 1 µl of an *Nostoc punctiforme ATCC 29133* DNA (prepared as described above)
- 0.25 mM dNTPs
- 0.2 mM NP196-1 (SEQ ID NO. 54)
- 0.2 mM NP196-2 (SEQ ID NO. 55)
- 5 µl of 10X PCR buffer (TAKARA)
- 0.25 µl of R Taq polymerase (TAKARA)
- 25.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
35X	94°C for 1 minute
	55°C for 1 minute

72°C for 3 minutes
1X 72°C for 10 minutes

PCR amplification with SEQ ID NO. 54 and SEQ ID NO. 55 resulted in a 792 bp fragment encoding a protein consisting of the entire primary sequence (NP196, SEQ ID NO. 56). Using standard methods, the amplicon was cloned into the PCR cloning vector pCR 2.1 (Invitrogen), producing the clone pNP196.

Sequencing of the pNP196 clone with the M13F and M13R primers confirmed a sequence which is identical to the DNA sequence from 140,571-139,810 of the database entry NZ_AABC01000196 (inversely oriented to the published database entry), except that G in position 140,571 was replaced by A in order to generate a standard ATG start codon. This nucleotide sequence was reproduced in an independent amplification experiment and thus represents the nucleotide sequence in the *Nostoc punctiforme ATCC 29133* used.

This clone, pNP196, was therefore used for cloning into the expression vector pJIT117 (Guerineau et al. 1988, Nucl. Acids Res. 16: 11380).

pJIT117 was modified by replacing the 35S terminator by the OCS terminator (octopine synthase) of the Agrobacterium tumefaciens Ti plasmid pTi15955 (database entry X00493, from position 12,541-12,350, Gielen et al. (1984) EMBO J. 3 835-846).

The DNA fragment which comprises the OCT terminator region was prepared by means of PCR using the plasmid pHELLSGATE (database entry AJ311874, Wesley et al. (2001) Plant J. 27 581-590, isolated from *E. coli* by standard methods) and the primers OCS-1 (SEQ ID NO. 58) and OCS-2 (SEQ ID NO. 59).

The PCR conditions were as follows:

The PCR for amplifying the DNA comprising the octopine synthase (OCS) terminator region (SEQ ID NO. 60) was carried out in a 50 µl reaction mixture containing:

- 1 ng of pHELLSGATE plasmid DNA
- 0.25 mM dNTPs
- 0.2 mM OCS-1 (SEQ ID NO. 58)
- 0.2 mM OCS-2 (SEQ ID NO. 59)
- 5 µl of 10X PCR buffer (Stratagene)
- 0.25 µl of Pfu polymerase (Stratagene)
- 28.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
35X	94°C for 1 minute
	50°C for 1 minute
	72°C for 1 minute
1X	72°C for 10 minutes

The 210 bp amplicon was cloned into the PCR cloning vector pCR 2.1 (Invitrogen) by using standard methods, producing the plasmid pOCS.

Sequencing of the pOCS clone confirmed a sequence which corresponds to a sequence section on the Agrobacterium tumefaciens Ti plasmid pTi15955 (database entry X00493), from position 12,541 to position 12,350.

Cloning was carried out by isolating the 210 bp SalI-XhoI fragment from pOCS and ligating it into the SalI-XhoI-cut pJIT117 vector.
This clone is referred to as pJO and was therefore used for cloning into the expression vector pJONP196.

Cloning was carried out by isolating the 782 bp SphI fragment from pNP196 and ligating it into the SphI-cut pJO vector. The clone which contains the *Nostoc punctiforme* NP196 ketolase in the correct orientation as an N-terminal translational fusion with the rbcS transit peptide is referred to as pJONP196.

Example 7:
Preparation of expression vectors for constitutive expression of NP196 ketolase from Nostoc *punctiforme ATCC 29133* in *Lycopersicon esculentum* and *Tagetes erecta*.

The *Nostoc punctiforme* NP196 ketolase is expressed in *L. esculentum* and in *Tagetes erecta* under the control of the constitutive promoter FNR (ferredoxin NADPH oxidoreductase, database entry AB011474, positions 70127 to 69493; WO03/006660) from *Arabidopsis thaliana*. The FNR gene starts at base pair 69492 and is annotated with "ferredoxin-NADP+ reductase". Expression was carried out using the pea transit peptide rbcS (Anderson et al. 1986, Biochem J. 240: 709-715).

The DNA fragment comprising the *Arabidopsis thaliana* FNR promoter region was prepared by means of PCR using genomic DNA (isolated from *Arabidopsis thaliana* by standard methods)

and the primers FNR-1 (SEQ ID NO. 61) and FNR-2 (SEQ ID NO. 62).

The PCR conditions were as follows:

The PCR for amplifying the DNA comprising the FNR promoter fragment FNR (SEQ ID NO. 63) was carried out in a 50 µl reaction mixture containing:

- 100 ng of *A. thaliana* genomic DNA
- 0.25 mM dNTPs
- 0.2 mM FNR-1 (SEQ ID NO. 61)
- 0.2 mM FNR-2 (SEQ ID NO. 62)
- 5 µl of 10X PCR buffer (Stratagene)
- 0.25 µl of Pfu polymerase (Stratagene)
- 28.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
35X	94°C for 1 minute
	50°C for 1 minute
	72°C for 1 minute
1X	72°C for 10 minutes

The 652 bp amplicon was cloned into the PCR cloning vector pCR 2.1 (Invitrogen) by using standard methods, producing the plasmid pFNR.

Sequencing of the pFNR clone confirmed a sequence which corresponds to a sequence section on chromosome 5 of Arabidopsis thaliana (database entry AB011474), from position 70127 to position 69493.

This clone is referred to as pFNR and was therefore used for cloning into the expression vector pJONP196 (described in example 6).

Cloning was carried out by isolating the 644 bp SmaI-HindIII fragment from pFNR and ligating it into the Ecl136II-HindIII-cut pJONP196 vector. The clone which contains the FNR promoter instead of the original d35S promoter and the fragment NP196 in the correct orientation as an N-terminal fusion with the rbcS transit peptide is referred to as pJOFNR:NP196.

An expression cassette for Agrobacterium-mediated transformation of the *Nostoc* NP196

ketolase into *L. esculentum* was prepared using the binary vector pSUN3 (WO02/00900).

The expression vector MSP105 was prepared by ligating the 1839 bp EcoRI-XhoI fragment from pJOFNR:NP196 with the EcoRI-XhoI-cut pSUN3 vector (figure 7, construct map). In figure 7, the fragment *FNR promoter* comprises the FNR promoter (635 bp), the fragment *rbcS TP FRAGMENT* comprises the pea rbcS transit peptide (194 bp), the fragment *NP196 KETO CDS* (761 bp) coding for *Nostoc punctiforme* NP196 ketolase, and the fragment *OCS terminator* (192 bp) comprises the polyadenylation signal of octopine synthase.

An expression cassette for *Agrobacterium*-mediated transformation of the expression vector containing the *Nostoc punctiforme* NP196 ketolase into Tagetes erecta was prepared using the binary vector pSUN5 (WO02/00900).

The Tagetes expression vector MSP106 was prepared by ligating the 1839 bp EcoRI-XhoI fragment from pJOFNR:NP196 with the EcoRI-XhoI-cut pSUN5 vector (figure 8, construct map). In figure 8, the fragment *FNR promoter* comprises the FNR promoter (635 bp), the fragment *rbcS TP FRAGMENT* comprises the pea rbcS transit peptide (194 bp), the fragment *NP196 KETO CDS* (761 bp) coding for *Nostoc punctiforme* NP196 ketolase, and the fragment *OCS terminator* (192 bp) comprises the polyadenylation signal of octopine synthase.

Example 8:
Preparation of expression vectors for flower-specific expression of NP196 ketolase from Nostoc *punctiforme ATCC 29133* in *Lycopersicon esculentum* and *Tagetes erecta*

Nostoc *punctiforme* NP196 ketolase was expressed in L. esculentum and Tagetes erecta by using the pea transit peptide rbcS (Anderson et al. 1986, Biochem J. 240: 709-715). Expression was carried out under the control of the flower-specific EPSPS promoter from Petunia hybrida (database entry M37029; nucleotide region 7-1787; Benfey et al. (1990) Plant Cell 2: 849-856).

The DNA fragment comprising the Petunia hybrida EPSPS promoter region (SEQ ID NO. 66) was prepared by means of PCR using genomic DNA (isolated from Petunia hybrida by standard methods) and the primers EPSPS-1 (SEQ ID NO. 64) and EPSPS-2 (SEQ ID NO. 65).

The PCR conditions were as follows:

The PCR for amplifying the DNA comprising the EPSPS promoter fragment (database entry M37029: nucleotide region 7-1787) was carried out in a 50 µl reaction mixture containing:

- 100 ng of A. thaliana genomic DNA

- 0.25 mM dNTPs
- 0.2 mM EPSPS-1 (SEQ ID NO. 64)
- 0.2 mM EPSPS-2 (SEQ ID NO. 65)
- 5 µl of 10X PCR buffer (Stratagene)
- 0.25 µl of Pfu polymerase (Stratagene)
- 28.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
35X	94°C for 1 minute
	50°C for 1 minute
	72°C for 2 minutes
1X	72°C for 10 minutes

The 1773 bp amplicon was cloned into the PCR cloning vector pCR 2.1 (Invitrogen) by using standard methods, producing the plasmid pEPSPS.

Sequencing of the pEPSPS clone confirmed a sequence which differs from the published EPSPS sequence (database entry M37029: nucleotide region 7-1787) only by two deletions (bases ctaagtttcagga at positions 46-58 of the sequence M37029; bases aaaaatat at positions 1422-1429 of the sequence M37029) and the base substitutions (T for G at position 1447 of the sequence M37029; A for C at position 1525 of the sequence M37029; A for G at position 1627 of the sequence M37029). The two deletions and the two base substitutions at positions 1447 and 1627 of the sequence M37029 were reproduced in an independent amplification experiment and thus represent the actual nucleotide sequence in the Petunia hybrida plants used.

The pEPSPS clone was therefore used for cloning into the expression vector pJONP196 (described in example 6).

Cloning was carried out by isolating the 1763 bp SacI-HindIII fragment from pEPSPS and ligating it into the SacI-HindIII-cut pJ0NP196 vector. The clone which contains the EPSPS promoter instead of the original d35S promoter is referred to as pJOESP:NP196. This expression cassette contains the NP196 fragment in the correct orientation as an N-terminal fusion with the rbcS transit peptide.

An expression vector for Agrobacterium-mediated transformation of the EPSPS-controlled Nostoc punctiforme ATCC 29133 NP196 ketolase into *L. esculentum* was prepared using the binary vector pSUN3 (WO02/00900).

The expression vector MSP107 was prepared by ligating the 2961 kbp SachI-XhoI fragment from pJOESP:NP196 with the SacI-XhoI-cut pSUN3 vector (figure 9, construct map). In figure 9, the fragment EPSPS comprises the EPSPS promoter (1761 bp), the fragment *rbcS TP FRAGMENT* comprises the pea rbcS transit peptide (194 bp), the fragment *NP196 KETO CDS* (761 bp) coding for *Nostoc punctiforme* NP196 ketolase, and the fragment *OCS terminator* (192 bp) comprises the polyadenylation signal of octopine synthase.

An expression vector for Agrobacterium-mediated transformation of the EPSPS-controlled Nostoc punctiforme NP196 ketolase into Tagetes erecta was prepared using the binary vector pSUN5 (WO02/00900).

The expression vector MSP108 was prepared by ligating the 2961 kbp SachI-XhoI fragment from pJOESP:NP196 with the SacI-XhoI-cut pSUN5 vector (figure 10, construct map). In figure 10, the fragment EPSPS comprises the EPSPS promoter (1761 bp), the fragment *rbcS TP FRAGMENT* comprises the pea rbcS transit peptide (194 bp), the fragment *NP196 KETO CDS* (761 bp) coding for *Nostoc punctiforme* NP196 ketolase, and the fragment *OCS terminator* (192 bp) comprises the polyadenylation signal of octopine synthase.

Example 9:
Amplification of a DNA encoding the entire primary sequence of NP195 ketolase from *Nostoc punctiforme ATCC 29133*

The DNA coding for NP195 ketolase from *Nostoc punctiforme ATCC 29133* was amplified from *Nostoc punctiforme ATCC 29133* (strain of the American Type Culture Collection) by means of PCR. The preparation of genomic DNA from a *Nostoc punctiforme ATCC 29133* suspension culture has been described in example 19.

The nucleic acid encoding a *Nostoc punctiforme ATCC 29133* ketolase was amplified from *Nostoc punctiforme ATCC 29133* by means of polymerase chain reaction (PCR) using a sense-specific primer (NP195-1, SEQ ID NO. 67) and an antisense-specific primer (NP195-2 SEQ ID NO. 68).

The PCR conditions were as follows:

The PCR for amplifying the DNA encoding a ketolase protein consisting of the entire primary sequence was carried out in a 50 µl reaction mixture which contained:

- 1 µl of a *Nostoc punctiforme ATCC 29133* DNA (prepared as described above)

- 0.25 mM dNTPs
- 0.2 mM NP195-1 (SEQ ID NO. 67)
- 0.2 mM NP195-2 (SEQ ID NO. 68)
- 5 µl of 10X PCR buffer (TAKARA)
- 0.25 µl of R Taq polymerase (TAKARA)
- 25.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
35X	94°C for 1 minute
	55°C for 1 minute
	72°C for 3 minutes
1X	72°C for 10 minutes

PCR amplification with SEQ ID NO. 67 and SEQ ID NO. 68 resulted in an 819 bp fragment encoding a protein consisting of the entire primary sequence (NP195, SEQ ID NO. 69). Using standard methods, the amplicon was cloned into the PCR cloning vector pCR 2.1 (Invitrogen), producing the clone pNP195.

Sequencing of the pNP195 clone with the M13F and M13R primers confirmed a sequence which is identical to the DNA sequence from 55,604-56,392 of the database entry NZ_AABC010001965, except that T at position 55,604 has been replaced with A in order to generate a standard ATG start codon. This nucleotide sequence was reproduced in an independent amplification experiment and thus represents the nucleotide sequence in the *Nostoc punctiforme ATCC 29133* used.

This clone, pNP195, was therefore used for cloning into the expression vector pJO (described in example 6). Cloning was carried out by isolating the 809 bp SphI fragment from pNP195 and ligating it into the SphI-cut pJO vector. The clone which contains the *Nostoc punctiforme* NP195 ketolase in the correct orientation as an N-terminal translation of fusion with the rbcS transit peptide is referred to as pJONP195.

Example 10:

Preparation of expression vectors for constitutive expression of NP195 ketolase from *Nostoc punctiforme ATCC 29133* in *Lycopersicon esculentum* and *Tagetes erecta*.

The *Nostoc punctiforme* NP195 ketolase is expressed in *L. esculentum* and in *Tagetes erecta* under the control of the constitutive promoter FNR (ferredoxin NADPH oxidoreductase,

database entry AB011474, positions 70127 to 69493; WO03/006660) from *Arabidopsis thaliana*. The FNR gene starts at base pair 69492 and is annotated with "ferredoxin-NADP+ reductase". Expression was carried out using the pea transit peptide rbcS (Anderson et al. 1986, Biochem J. 240: 709-715).

The clone pFNR (described in example 7) was therefore used for cloning into the expression vector pJONP195 (described in example 10).

Cloning was carried out by isolating the 644 bp Sma-HindIII fragment from pFNR and ligating it into the EcI136II-HindIII-cut pJONP195 vector. The clone which contains the FNR promoter instead of the original d35S promoter and the fragment NP195 in the correct orientation as an N-terminal fusion with the rbcS transit peptide is referred to as pJOFNR:NP195.

An expression cassette for Agrobacterium-mediated transformation of the *Nostoc punctiforme* NP195 ketolase in *L. esculentum* was prepared using the binary vector pSUN3 (WO02/00900).

The expression vector MSP109 was prepared by ligating the 1866 bp EcoRI-XhoI fragment from pJOFNR:NP195 with the EcoRI-XhoI-cut pSUN3 vector (figure 11, construct map). In figure 11, the fragment *FNR promoter* comprises the FNR promoter (635 bp), the fragment *rbcS TP FRAGMENT* comprises the pea rbcS transit peptide (194 bp), the fragment *NP195 KETO CDS* (789 bp) coding for *Nostoc punctiforme* NP195 ketolase, and the fragment *OCS terminator* (192 bp) comprises the polyadenylation signal of octopine synthase.

An expression cassette for *Agrobacterium*-mediated transformation of the expression vector containing the Nostoc punctiforme punctiforme NP195 ketolase into Tagetes erecta was prepared using the binary vector pSUN5 (WO02/00900).

The Tagetes expression vector MSP110 was prepared by ligating the 1866 bp EcoRI-XhoI fragment from pJOFNR:NP195 with the EcoRI-XhoI-cut pSUN5 vector (figure 12, construct map). In figure 12, the fragment *FNR promoter* comprises the FNR promoter (635 bp), the fragment rbcS TP FRAGMENT comprises the pea rbcS transit peptide (194 bp), the fragment NP195 KETO CDS (789 bp) coding for Nostoc punctiforme NP195 ketolase, and the fragment OCS terminator (192 bp) comprises the polyadenylation signal of octopine synthase.

Example 11:

Preparation of expression vectors for flower-specific expression of NP195 ketolase from Nostoc punctiforme ATCC 29133 in Lycopersicon esculentum and Tagetes erecta

Nostoc punctiforme NP195 ketolase was expressed in L. esculentum and Tagetes erecta by

using the pea transit peptide rbcS (Anderson et al. 1986, Biochem J. 240: 709-715). Expression was carried out under the control of the flower-specific EPSPS promoter from Petunia hybrida (database entry M37029; nucleotide region 7-1787; Benfey et al. (1990) Plant Cell 2: 849-856).

The pEPSPS clone (described in example 8) was therefore used for cloning into the expression vector pJONP195 (described in example 10).

Cloning was carried out by isolating the 1763 bp SacI-HindIII fragment from pEPSPS and ligating it into the SacI-HindIII-cut pJONP195 vector. The clone which contains the EPSPS promoter instead of the original d35S promoter is referred to as pJOESP:NP195. This expression cassette contains the NP195 fragment in the correct orientation as an N-terminal fusion with the rbcS transit peptide.

An expression vector for Agrobacterium-mediated transformation of the EPSPS-controlled Nostoc punctiforme ATCC 29133 NP195 ketolase into *L. esculentum* was prepared using the binary vector pSUN3 (WO02/00900).

The expression vector MSP111 was prepared by ligating the 2988 kbp SachI-XhoI fragment from pJOESP:NP195 with the SacI-XhoI-cut pSUN3 vector (figure 13, construct map). In figure 13, the fragment EPSPS comprises the EPSPS promoter (1761 bp), the fragment *rbcS TP FRAGMENT* comprises the pea rbcS transit peptide (194 bp), the fragment *NP195 KETO CDS* (789 bp) coding for *Nostoc punctiforme* NP195 ketolase, and the fragment *OCS terminator* (192 bp) comprises the polyadenylation signal of octopine synthase.

An expression vector for Agrobacterium-mediated transformation of the EPSPS-controlled Nostoc punctiforme NP195 ketolase into Tagetes erecta was prepared using the binary vector pSUN5 (WO02/00900).

The expression vector MSP112 was prepared by ligating the 2988 kbp SachI-XhoI fragment from pJOESP:NP195 with the SacI-XhoI-cut pSUN5 vector (figure 14, construct map). In figure 14, the fragment EPSPS comprises the EPSPS promoter (1761 bp), the fragment *rbcS TP FRAGMENT* comprises the pea rbcS transit peptide (194 bp), the fragment *NP195 KETO CDS* (789 bp) coding for *Nostoc punctiforme* NP195 ketolase, and the fragment *OCS terminator* (192 bp) comprises the polyadenylation signal of octopine synthase.

Example 12:

Amplification of a DNA encoding the entire primary sequence of NODK ketolase from *Nodularia spumignea NSOR10*

The DNA coding for the *Nodularia spumignea NSOR10* ketolase was amplified from *Nodularia spumignea NSOR10* by means of PCR.

To prepare genomic DNA from a *Nodularia spumignea NSOR10* suspension culture which had grown in *BG 11* medium (1.5 g/l $NaNO_3$, 0.04 g/l $K_2PO_4x3H_2O$, 0.075 g/l $MgSO_4xH_2O$, 0.036 g/l $CaCl_2x2H_2O$, 0.006 g/l citric acid, 0.006 g/l ferric ammonium citrate, 0.001 g/l EDTA disodium magnesium, 0.04 g/l Na_2CO_3, 1 ml of trace metal mix A5+Co (2.86 g/l H_3BO_3, 1.81 g/l $MnCl_2x4H_2O$, 0.222 g/l $ZnSO_4x7H_2O$, 0.39 g/l $NaMoO_4x2H_2O$, 0.079 g/l $CuSO_4x5H_2O$, 0.0494 g/l $Co(NO_3)_2x6H_2O$) at 25°C with constant shaking (150 rpm) and under continuous light for 1 week, the cells were harvested by centrifugation, frozen in liquid nitrogen and ground to a powder in a mortar.

Protocol for isolating DNA from *Nodularia spumignea NSOR10*:

The bacterial cells were pelleted from a 10 ml liquid culture by centrifugation at 8000 rpm for 10 minutes. The bacterial cells were then crushed and ground in liquid nitrogen, using a mortar. The cell material was resuspended in 1 ml of 10 mM Tris HCl (pH 7.5) and transferred to an Eppendorf reaction vessel (volume: 2 ml). After addition of 100 µl of proteinase K (concentration: 20 mg/ml), the cell suspension was incubated at 37°C for 3 hours. The suspension was then extracted with 500 µl of phenol. After centrifugation at 13 000 rpm for 5 minutes, the upper, aqueous phase was transferred to a new 2 ml Eppendorf reaction vessel. Extraction with phenol was repeated 3 times. The DNA was precipitated by adding 1/10 volume of 3 M sodium acetate (pH 5.2) and 0.6 volume of isopropanol and then washed with 70% ethanol. The DNA pellet was dried at room temperature, taken up in 25 µl of water and dissolved with heating to 65°C.

The nucleic acid encoding a *Nodularia spumignea NSOR10* ketolase was amplified from *Nodularia spumignea NSOR10* by means of polymerase chain reaction (PCR) using a sense-specific primer (NODK-1, SEQ ID NO. 71) and an antisense-specific primer (NODK-2 SEQ ID NO. 72).

The PCR conditions were as follows:

The PCR for amplifying the DNA encoding a ketolase protein consisting of the entire primary sequence was carried out in a 50 µl reaction mixture which contained:

- 1 µl of a *Nodularia spumignea NSOR10* DNA (prepared as described above)
- 0.25 mM dNTPs
- 0.2 mM NODK-1 (SEQ ID NO. 71)
- 0.2 mM NODK-2 (SEQ ID NO. 72)

- 5 µl of 10X PCR buffer (TAKARA)
- 0.25 µl of R Taq polymerase (TAKARA)
- 25.8 µl of distilled water

The PCR was carried out under the following cycle conditions:

1X	94°C for 2 minutes
35X	94°C for 1 minute
	55°C for 1 minute
	72°C for 3 minutes
1X	72°C for 10 minutes

PCR amplification with SEQ ID NO. 71 and SEQ ID NO. 72 resulted in a 720 bp fragment encoding a protein consisting of the entire primary sequence (NODK, SEQ ID NO. 73). Using standard methods, the amplicon was cloned into the PCR cloning vector pCR 2.1 (Invitrogen), producing the clone pNODK.

Sequencing of the pNODK clone with the M13F and M13R primers confirmed a sequence which is identical to the DNA sequence from 2130-2819 of the database entry AY210783 (inversely oriented to the published database entry). This nucleotide sequence was reproduced in an independent amplification experiment and thus represents the nucleotide sequence in the *Nodularia spumignea NSOR10* used.

This clone, pNODK, was therefore used for cloning into the expression vector pJO (described in example 6). Cloning was carried out by isolating the 710 bp SphI fragment from pNODK and ligating it into the SphI-cut pJO vector. The clone which contains the *Nodularia spumignea* NODK ketolase in the correct orientation as an N-terminal translational fusion with the rbcS transit peptide is referred to as pJONODK.

Example 13:
Preparation of expression vectors for constitutive expression of NODK ketolase from *Nodularia spumignea NSOR10* in *Lycopersicon esculentum* and *Tagetes erecta*

The *Nodularia spumignea NSOR10* NODK ketolase was expressed in *L. esculentum* and in *Tagetes erecta* under the control of the constitutive promoter FNR (ferredoxin NADPH oxidoreductase, database entry AB011474, positions 70127 to 69493; WO03/006660) from *Arabidopsis thaliana*. The FNR gene starts at base pair 69492 and is annotated with "ferredoxin-NADP+ reductase". Expression was carried out using the pea transit peptide rbcS (Anderson et al. 1986, Biochem J. 240: 709-715).

The clone pFNR (described in example 7) was therefore used for cloning into the expression vector pJONODK (described in example 12).

Cloning was carried out by isolating the 644 bp Sma-HindIII fragment from pFNR and ligating it into the EcI136II-HindIII-cut pJONODK vector. The clone which contains the FNR promoter instead of the original d35S promoter and the fragment NODK in the correct orientation as an N-terminal fusion with the rbcS transit peptide is referred to as pJOFNR:NODK.

An expression cassette for Agrobacterium-mediated transformation of the *Nodularia spumignea NSOR10* NODK ketolase into *L. esculentum* was prepared using the binary vector pSUN3 (WO02/00900).

The expression vector MSP113 was prepared by ligating the 1767 bp EcoRI-XhoI fragment from pJOFNR:NODK with the EcoRI-XhoI-cut pSUN3 vector (figure 15, construct map). In figure 15, the fragment *FNR promoter* comprises the FNR promoter (635 bp), the fragment *rbcS TP FRAGMENT* comprises the pea rbcS transit peptide (194 bp), the fragment *NODK KETO CDS* (690 bp) coding for *Nodularia spumignea NSOR10* NODK ketolase, and the fragment *OCS terminator* (192 bp) comprises the polyadenylation signal of octopine synthase.

An expression cassette for *Agrobacterium*-mediated transformation of the expression vector containing the *Nodularia spumignea NSOR10* punctiforme NODK ketolase into Tagetes erecta was prepared using the binary vector pSUN5 (WO02/00900).

The Tagetes expression vector MSP114 was prepared by ligating the 1767 bp EcoRI-XhoI fragment from pJOFNR:NODK with the EcoRI-XhoI-cut pSUN5 vector (figure 16, construct map). In figure 16, the fragment *FNR promoter* comprises the FNR promoter (635 bp), the fragment rbcS TP FRAGMENT comprises the pea rbcS transit peptide (194 bp), the fragment NODK KETO CDS (690 bp) coding for *Nodularia spumignea NSOR10* NODK ketolase, and the fragment OCS terminator (192 bp) comprises the polyadenylation signal of octopine synthase.

Example 14:
Preparation of expression vectors for flower-specific expression of NODK ketolase from *Nodularia spumignea NSOR10* in *Lycopersicon esculentum* and *Tagetes erecta*

Nodularia spumignea NSOR10 NODK ketolase was expressed in L. esculentum and Tagetes erecta by using the pea transit peptide rbcS (Anderson et al. 1986, Biochem J. 240: 709-715). Expression was carried out under the control of the flower-specific EPSPS promoter from Petunia hybrida (database entry M37029; nucleotide region 7-1787; Benfey et al. (1990) Plant

Cell 2: 849-856).

The pEPSPS clone (described in example 8) was therefore used for cloning into the expression vector pJONODK (described in example 12).

Cloning was carried out by isolating the 1763 bp SacI-HindIII fragment from pEPSPS and ligating it into the SacI-HindIII-cut pJONODK vector. The clone which contains the EPSPS promoter instead of the original d35S promoter is referred to as pJOESP:NODK. This expression cassette contains the NODK fragment in the correct orientation as an N-terminal fusion with the rbcS transit peptide.

An expression vector for Agrobacterium-mediated transformation of the EPSPS-controlled *Nodularia spumignea NSOR10* NODK ketolase into *L. esculentum* was prepared using the binary vector pSUN3 (WO02/00900).

The expression vector MSP115 was prepared by ligating the 2889 kbp SachI-XhoI fragment from pJOESP:NODK with the SacI-XhoI-cut pSUN3 vector (figure 17, construct map). In figure 17, the fragment EPSPS comprises the EPSPS promoter (1761 bp), the fragment *rbcS TP FRAGMENT* comprises the pea rbcS transit peptide (194 bp), the fragment *NODK KETO CDS* (690 bp) coding for *Nodularia spumignea NSOR10* NODK ketolase, and the fragment *OCS terminator* (192 bp) comprises the polyadenylation signal of octopine synthase.

An expression vector for Agrobacterium-mediated transformation of the EPSPS-controlled *Nodularia spumignea NSOR10* NODK ketolase into Tagetes erecta was prepared using the binary vector pSUN5 (WO02/00900).

The expression vector MSP116 was prepared by ligating the 2889 kbp SachI-XhoI fragment from pJOESP:NODK with the SacI-XhoI-cut pSUN5 vector (figure 18, construct map). In figure 18, the fragment EPSPS comprises the EPSPS promoter (1761 bp), the fragment *rbcS TP FRAGMENT* comprises the pea rbcS transit peptide (194 bp), the fragment *NODK KETO CDS* (690 bp) coding for *Nodularia spumignea NSOR10* NODK ketolase, and the fragment *OCS terminator* (192 bp) comprises the polyadenylation signal of octopine synthase.

Example 15:

Preparation of transgenic Lycopersicon esculentum plants

Tomato plants were transformed and regenerated according to the published method by Ling and coworkers (Plant Cell Reports (1998), 17: 843-847). In the case of the variety Microtom, selection was carried out using a higher kanamycin concentration (100 mg/l).

Cotyledons and hypocotyls of seedlings seven to ten days old of the Microtom line were used as initial explant for the transformation. The culture medium of Murashige and Skoog (1962: Murashige and Skoog, 1962, Physiol. Plant 15, 473-) containing 2% sucrose, pH 6.1 was used for germination. Germination took place at 21°C with low light (20-100 µE). After seven to ten days, the cotyledons were divided transversely, and the hypocotyls were cut into sections about 5-10 mm long and placed on the MSBN medium (MS, pH 6.1, 3% sucrose + 1 mg/l BAP, 0.1 mg/l NAA) which had been charged the previous day with suspension-cultivated tomato cells. The tomato cells were covered, free of air bubbles, with sterile filter paper. The explants were precultured on the described medium for three to five days. Cells of the Agrobacterium tumefaciens strain LBA4404 were transformed individually with the plasmids pS3FNR:NOST, pS3AP3:NOST, pS3FNR:NP196, pS3EPS:NP196, pS3FNR:NP195, pS3EPS:NP195, pS3FNR:NODK and pS3EPS:NODK. In each case, an overnight culture of the individual Agrobacterium strains transformed with the binary vectors pS3FNRNOST, pS3AP3NOST, pS3FNR:NP196, pS3EPS:NP196, pS3FNR:NP195, pS3EPS:NP195, pS3FNR:NODK and pS3EPS:NODK was cultured in YEB medium with kanamycin (20 mg/l) at 28°C, and the cells were centrifuged. The bacterial pellet was resuspended in liquid MS medium (3% sucrose, pH 6.1) and adjusted to an optical density of 0.3 (at 600 nm). The precultured explants were transferred into the suspension and incubated at room temperature with gentle shaking for 30 minutes. Subsequently, the explants were dried with sterile filter paper and returned to their preculture medium for the three-day coculture (21°C).

After the coculture, the explants were transferred to MSZ2 medium (MS pH 6.1 + 3% sucrose, 2 mg/l zeatin, 100 mg/l kanamycin, 160 mg/l timentin) and stored for the selective regeneration at 21°C under weak light conditions (20-100 µE, light rhythm 16 h/8 h). The explants were transferred every two to three weeks until shoots formed. It was possible to detach small shoots from the explant and root them on MS (pH 6.1 + 3% sucrose), 160 mg/l timentin, 30 mg/l kanamycin, 0.1 mg/l IAA. Rooted plants were transferred into the glasshouse.

According to the transformation method described above, the following lines were obtained with the expression constructs below:

obtained with pS3FNR:NOST were: MSP101-1, MSP101-2, MSP101-3

obtained with pS3AP3:NOST were: MSP103-1, MSP103-2, MSP103-3

obtained with pS3FNR:NP196 were: MSP105-1, MSP105-2, MSP105-3

obtained with pS3EPS:NP196 were: MSP107-1, MSP107-2, MSP107-3

obtained with pS3FNR:NP195 were: MSP109-1, MSP109-2, MSP109-3

obtained with pS3EPS:NP195 were: MSP111-1, MSP111-2, MSP111-3

obtained with pS3FNR:NODK were: MSP113-1, MSP113-2, MSP113-3

obtained with pS3EPS:NODK were: MSP115-1, MSP115-2, MSP115-3

Example 16:
Preparation of transgenic Tagetes plants

Tagetes seeds are sterilized and placed on germination medium (MS medium; Murashige and Skoog, Physiol. Plant. 15 (1962), 473-497) pH 5.8, 2% sucrose). Germination takes place in a temperature/light/time interval of 18-28°C/20-200 µE/3-16 weeks, but preferably at 21°C, 20-70 µE, for 4-8 weeks.

All leaves of the plants which have developed in vitro by then are harvested and cut transverse to the middle. The leaf explants resulting therefrom, with a size of 10-60 mm^2, are stored during the preparation in liquid MS medium at room temperature for not more than 2 h.

Any Agrobacterium tumefaciens strain, but preferably a supervirulent strain such as, for example, EHA105 with an appropriate binary plasmid which may harbor a selection marker gene (preferably *bar* or *pat*) and one or more trait or reporter genes (pS5FNR:NOST, pS5AP3:NOST, pS5FNR:NP196, pS5EPS:NP196, pS5FNR:NP195, pS5EPS:NP195, pS5FNR:NODK and pS5EPS:NODK) cultivated overnight and used for cocultivation with the leaf material. The bacterial strain can be cultured as follows: a single colony of the appropriate strain is inoculated in YEB (0.1% yeast extract, 0.5% beef extract, 0.5% peptone, 0.5% sucrose, 0.5% magnesium sulfate x 7 H_2O) with 25 mg/l kanamycin and cultured at 28°C for 16 to 20 h. The bacterial suspension is then harvested by centrifugation at 6000 g for 10 min and resuspended in liquid MS medium so as to result in an OD_{600} of approx. 0.1 to 0.8. This suspension is used for cocultivation with the leaf material.

Immediately before the cocultivation, the MS medium in which the leaves have been stored is replaced with the bacterial suspension. Incubation of the leaves in the agrobacterial suspension took place at room temperature with gentle shaking for 30 min. The infected explants are then put on an MS medium solidified with agar (e.g. 0.8% Plant Agar (Duchefa, NL)), with growth regulators such as, for example, 3 mg/l benzylaminopurin (BAP) and 1 mg/l indolylacetic acid (IAA). The orientation of the leaves on the medium is immaterial. Cultivation of the explants

takes place for 1 to 8 days, but preferably for 6 days, during which the following conditions can be applied: light intensity: 30-80 µmol/m^2 x sec, temperature: 22-24°C, 16/8 hours light/dark alternation. The cocultivated explants are then transferred to fresh MS medium, preferably with the same growth regulators, this second medium additionally containing an antibiotic to suppress bacterial growth. Timentin in a concentration of 200 to 500 mg/l is very suitable for this purpose. The second selective component employed is one for selecting for successful transformation. Phosphinothricin in a concentration of 1 to 5 mg/l selects very efficiently, but other selective components according to the method to be used are also conceivable.

After one to three weeks in each case, the explants are transferred to fresh medium until plumules and small shoots develop, which are then transferred to the same basal medium including timentin and PPT or alternative components with growth regulators, namely, for example, 0.5 mg/l indolylbutyric acid (IBA) and 0.5 mg/l gibberillic acid GA_3, for rooting. Rooted shoots can be transferred into the glasshouse.

In addition to the method described, the following advantageous modifications are possible:

Before the explants are infected with the bacteria, they may be preincubated on the medium described above for cocultivation for 1 to 12 days, preferably 3-4. This is followed by infection, cocultivation and selective regeneration as described above.

The pH for regeneration (normally 5.8) may be lowered to pH 5.2. This improves control of agrobacterial growth.

Addition of $AgNO_3$ (3-10 mg/l) to the regeneration medium improves the condition of the culture, including the regeneration itself.

Components which reduce phenol formation and are known to the skilled worker, such as, for example, citric acid, ascorbic acid, PVP and many others, have beneficial effects on the culture.

Liquid culture medium may also be used for the whole method. The culture may also be incubated on commercially available supports which are positioned on the liquid medium.

According to the above-described transformation method, the following lines were obtained with the expression constructs below:

obtained with pS5FNR:NOST were, for example: MSP102-1, MSP102-2, MSP102-3

obtained with pS5AP3:NOST were, for example: MSP104-1, MSP104-2, MSP104-3

obtained with pS5FNR:NP196 were: MSP106-1, MSP106-2, MSP106-3

obtained with pS5EPS:NP196 were: MSP108-1, MSP108-2, MSP108-3

obtained with pS5FNR:NP195 were: MSP110-1, MSP110-2, MSP110-3

obtained with pS5EPS:NP195 were: MSP112-1, MSP112-2, MSP112-3

obtained with pS5FNR:NODK were: MSP114-1, MSP114-2, MSP114-3

obtained with pS5EPS:NODK were: MSP116-1, MSP116-2, MSP116-3

Example 17

Characterization of the transgenic plant flowers

Example 9.1

Separation of carotenoid esters in petals of transgenic plants

General protocol:

The petals of the transgenic plants are ground in liquid nitrogen and the petal powder (about 40 mg) is extracted with 100% acetone (three times with 500 µl each). The solvent is evaporated and the carotenoids are resuspended in 100-200 µl of petroleum ether/acetone (5:1, v/v).

The carotenoids are fractionated in concentrated form by means of thin layer chromatography (TLC) on Silica60 F254 plates (Merck) in an organic solvent (petroleum ether/acetone; 5:1), according to their phobicity. Yellow (xanthophyll esters), red (ketocarotenoid esters) and orange bands (mixture of xanthophyll and ketocarotenoid esters) on the TLC are scraped out.

The carotenoids bound to silica are eluted three times with 500 µl of acetone, the solvent is evaporated and the carotenoids are fractionated and identified by means of HPLC.

It is possible to distinguish between mono- and diesters of carotenoids by means of a C30 reverse phase column. HPLC run conditions were virtually identical to those of a published method (Frazer et al. (2000), Plant Journal 24(4): 551-558). The following process conditions were set.

Separating column: Prontosil C30 column, 250 x 4.6 mm (Bischoff, Leonberg, Germany)

Flow rate: 1.0 ml/min

Eluents: eluent A - 100% methanol

Eluent B - 80% methanol, 0.2% ammonium acetate

Eluent C - 100% t-butyl methyl ether

Gradient profile:

Time	Flow rate	% eluent A	% eluent B	% eluent C
12.0	1.0	95.0	5.0	0
12.1	1.0	80.0	5.0	15.0
22.0	1.0	76.0	5.0	19.0
22.0	1.0	66.5	5.0	28.5
38.0	1.0	15.0	5.0	80.0
45.0	1.0	95.0	5.0	0
46.0	1.0	95.0	5.0	0
46.1	1.0	95.0	5.0	0

Detection: 300-500 nm

It is possible to identify carotenoids on the basis of UV-VIS spectra.

Petal material of the transgenic tomato plants is ground and extracted with acetone. Extracted carotenoids are fractionated by means of TLC. In the lines, mono- and diesters of keto-carotenoids may be detected; the monoesters are present at a distinctly lower concentration than the diesters.

Example 18

Enzymatic hydrolysis of carotenoid esters and identification of carotenoids

General protocol

Ground petal material (30-100 mg fresh weight) is extracted with 100% acetone (three times 500 µl; shaking for about 15 minutes each time). The solvent is evaporated. Carotenoids are then taken up in 495 µl of acetone and, after addition of 4.95 ml of potassium phosphate buffer (100 mM, pH 7.4), thoroughly mixed. This is followed by addition of about 17 mg of bile salts (Sigma) and 149 µl of an NaCl/$CaCl_2$ solution (3M NaCl and 75 mM $CaCl_2$). The suspension is incubated at 37°C for 30 minutes. For the enzymatic hydrolysis of the carotenoid esters, 595 µl of a lipase solution (50 mg/ml lipase type 7 from Candida rugosa (Sigma)) are added and incubated at 37°C with shaking. After about 21 hours, a further 595 µl of lipase are added, with renewed incubation at 37°C for at least 5 hours. Then about 700 mg of $Na_2SO_4x10H_2O$ are

dissolved in the solution. After addition of 1800 µl of petroleum ether, the carotenoids are extracted into the organic phase by vigorous mixing. This extraction is repeated until the organic phase remains colorless. The petroleum ether fractions are combined and the petroleum ether is evaporated. Free carotenoids are taken up in 100-120 µl of acetone. Free carotenoids can be identified on the basis of retention time and UV-VIS spectra by means of HPLC and a C30 reverse phase column.